ERIC Documents Resume	See Instruction Sheet	DATE OF RESUME	DATE RECEIVED

A. PUBLICATION DATA

1. ERIC NUMBER	2. ORIGINATOR CONTROL NO(S).	3. IDENTIFICATION NO.
ED001145		CR 1929

DO NOT WRITE IN SPACE BELOW

4. AUTHOR(S):
Kennedy, Wallace A. and Willcutt, Herman C.

5. TITLE OF DOCUMENT ABSTRACTED:
Motivation of School Children

6. PUBLICATION TITLE:
N.A.

7. EDITOR(S):
N.A.

8. PLACE OF ORIGIN
ORG. Florida State Univ. Human Dev. Clinic
ST. Tallahassee, Fla.
CITY, STATE
TELE. EXT.
INST/LOC CODE

9. PUBLISHER OR SOURCE
Cooperative Research, Office of Edu., Dept. of HEW, Washington, D. C.

10. VOL.	11. PAGES	12. MO./YR.	13. TOTAL PP.
N.A.	N.A.	Oct. 63	93

MICROFILM COPY AVAILABLE
YES ☒ NO ☐

COPYRIGHT GRANTED
YES ☐ NO ☒

COPYRIGHT RELEASE FILED
YES ☐ NO ☐

TYPE OF REPORT
PROPOSAL ☐
IN-PROGRESS ☐
FINAL ☒
FOLLOW-UP ☐

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

B. INDEXING DATA

14. SUBJECT AREA

15. TYPE OF ENDEAVOR

16. INDEXING TERMS (6-12)

motivation effects	elementary edu.	educational research	discrimination tasks	
reaction time	secondary edu.	intelligence effects	praise-blame effects	
classroom techniques	psychological research	age effects	socio-psychological variables	
		sex effects		

17. ABSTRACT (MAX. 250 WORDS) An extensive review of the relevant literature since 1900 seems to indicate that in some way there is an interaction between intelligence, age, social group, and the effects of praise and blame on school children. Inconsistencies in experimental factors have prevented the making of any definitive statements regarding the problem.

An investigation of the effects of praise and blame on a discrimination task under the variables of grade, intelligence, sex, social class, school and examiner was made. Seven hundred and twenty subjects were drawn from a large southern metropolitan area were divided into groups based upon three reward conditions, four grade levels, three levels of intelligence and two races. Although all of the subjects could solve the task used in this study, the speed with which it was solved was highly variable. The analysis of this variance was the major statistic of the study. Instruments used were the California Test of Mental Maturity to establish intelligence level; a Wollensak type recorder to monitor all tests; an Esterline Angus 20-pen recorder to record all subject and examiner responses; and a four-key discrimination box designed for the study. The 32 oddity-problem stimulus cards presented four patterns one of which was different from the others. The problem was to identify the odd pattern as quickly as possible. Between trials some subjects were given blame, some praise and told to do better, and the controls were asked questions of a biographical nature. All subjects given praise responded with decreased mean reaction time on trial two. Out of all the analyses of variance done the most striking finding was the consistency of the effects of blame and praise, irrespective of some rather basic differences in subject variables. The conclusion was that in situations involving perceptual motor skills, with speed as the criterion measure, reproof should have a debilitating effect upon performance while praise results in a somewhat higher increase in performance than that coming from practice alone.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION	20. DIVISION OR DEPT.	21. CONTRACT NO. / PROJECT NO.
	22. TOTAL COST OF PROJECT	
	23. DURATION OF PROJECT	U.S. Department of Health, Education, and Welfare Office of Education

ED001145

MOTIVATION OF SCHOOL CHILDREN

By Wallace A. Kennedy and Herman C. Willcutt

REPRINTED BY PERMISSION
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN PRINTED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

C.P.E. 1929

MOTIVATION OF SCHOOL CHILDREN

Wallace A. Kennedy and Herman C. Willcutt

THE FLORIDA STATE UNIVERSITY
Human Development Clinic
Tallahassee, Florida
October, 1963

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education, and Welfare.

C.R.P. 1929

Acknowledgements

The writers are indebted to Dr. Cranford Burns, Supervisor of the Board of the School Commission of Mobile County, and Mr. Floyd C. Replogle, Director of Child Guidance in Mobile County, for their assistance in carrying out the project in the Mobile County School Systems, and to the many teachers and principals who aided us.

Dr. F. J. King, of the Educational Research and Testing Department at Florida State University, provided invaluable assistance to the project, but especially in programming the complex repeated measures analysis of variance.

Dr. Ralph Witherspoon, Director of the Institute of Human Development, and the entire Institute staff gave us tremendous logistical support throughout the project, but especially in the pilot phase of the work, which was supported entirely in the Institute.

Table of Contents

Page

List of Tables viii

List of Figures x

Review of the Literature 1

Up to 1930

Binet & Vaschide (1897)

Kirby (1913)

Gilchrist (1916) and Gates & Rissland (1923)

Hurlock (1924)

Hurlock (1925a)

Hurlock (1925b)

Cohen (1927)

Bird (1927)

Hurlock (1931)

Summary up to 1930

1930-1940

Warden & Cohen (1931)

Davis & Ballard (1932)

Chase (1932)

Anderson & Smith (1933)

Brenner (1934)

Benton (1936)

Forlano & Axelrod (1937)

Murphy, Murphy, & Newcomb (1937)

Schmidt (1941)

Summary 1930 to 1940

Page

1940-1950
Potter (1943)
Klugman (1944)
Thompson & Hunnicutt (1944)
Summary 1940 to 1950
1950-1960
Mech, Kapos, Hurst, & Auble (1954)
Kent (1956)
Silverman (1957)
Terrell & Kennedy (1957)
Ellis & Distephano (1958)
Sandstrom & Weinz (1958)
Summary 1950 to 1960
1960 to Date
Dollins, Angelino, & Mech (1960)
Stevenson & Snyder (1960)
Fischer (1962)
Kennedy, Turner, & Lindner (1962)
Metz (1962)
Tiber & Kennedy (1963)
Willcutt & Kennedy (1963)
Summary 1960 to date
Over-all Summary

Problem 28

Method 29
Instruments
Stimulus Cards
Initial Procedure
Trials
Criterion Measure

Page

Statistical Design
General Population Characteristics

Results 40
White Sample
Summary for White Sample
Negro Sample
Summary for Negro Sample
Over-all Summary

Conclusions 60

Summary 63

References 73

Appendix 77

List of Tables

Table		Page
1.	Distribution of Occupations for Parent Population and Families of Children in the Sample	38
2.	Distribution of Educational Attainment of Parent Population and Parents of Children in the Sample	39
3.	Means and Standard Deviations of Subjects' Chronological Age in Months by Grade and Race	40
4.	White Sample Analysis of Variance Relationships between Reward, Trial, Grade, and Intelligence . .	41
5.	White Sample Mean Reaction Time by Trial	42
6.	White Sample Mean Reaction Time by Grade	42
7.	White Sample Mean Reaction Time by Intelligence . .	43
8.	White Sample Mean Reaction Time for Trials One and Two by Reward	44
9.	White Sample Mean Reaction Time for Trials One and Two by Reward and Grade	45
10.	White sample Mean Reaction Time for Trials One and Two by Reward, Grade, and Intelligence	46
11.	White Sample Analysis of Variance Relationships between Reward, Trial, Grade, and Sex	50
12.	White Sample Mean Reaction Time by Reward, Grade, and Sex	51

Table Page

13. White Sample Mean Reaction Time for Trials One and Two by Sex 52

14. Negro Sample Analysis of Variance Relationships between Reward, Trail, Grade, and Intelligence . . 54

15. Negro Sample Mean Reaction Time by Trial 55

16. Negro Sample Mean Reaction Time by Intelligence . . 55

17. Negro Sample Mean Reaction Time for Trials One and Two by Reward 56

18. Negro Sample Mean Reaction Time for Trials One and Two by Reward and Grade 57

19. Negro Sample Analysis of Variance Relationships between Reward, Trial, Grade, and Sex 58

List of Figures

Figure		Page
1.	Testing Complex	31
2.	Diagram of Apparatus Wiring	32
3.	Analysis of Variance Cell Division	37
4.	Stimulus Cards at Difficulty Level 1	78
5.	Stimulus Cards at Difficulty Level 2	79
6.	Stimulus Cards at Difficulty Level 3	80
7.	Stimulus Cards at Difficulty Level 4	81
8.	Stimulus Cards at Difficulty Level 5	82
9.	Stimulus Cards at Difficulty Level 6	83
10.	Stimulus Cards at Difficulty Level 7	84
11.	Stimulus Cards at Difficulty Level 8	85

MOTIVATION OF SCHOOL CHILDREN

Wallace A. Kennedy and Herman C. Willcutt
Florida State University

Experimental studies using various rewards and punishments for motivating discrimination, learning behavior, and performance in human and animal subjects compose the greater part of the literature of psychology. The history of child psychology is spanned by studies seeking to evaluate the effectiveness of praise and blame as incentives for school children, a literature sufficiently rich, broad, and confusing in and of itself to warrant special attention. The present review will be confined to empirical studies considering the verbal incentives praise and blame in connection with discrimination, learning, and motor tasks in school-age children below the college level. Only philosophical and theoretical papers directly derived from empirical data will be considered.

Review of the Literature

Up to 1930

Binet & Vaschide (1897)

When Binet & Vaschide (1897), eight years prior to Binet's monumental work on intelligence, used two trials on an ergograph and a hand dynamometer to measure the effect of verbal encouragement on the physical output of children, all 43 subjects improved their score. No mention was made, however, of a control group, the exact age of the children, or a warm-up period.

Kirby (1913)

In an effort to evaluate the effect of verbal encouragement, Kirby (1913) utilized 1350 third- and fourth-grade children working addition and division problems. The subjects were tested on Thorndike arithmetic tests, all given verbal encouragement to improve their own scores, knowledge of results, followed by a practice period, and then a retest. No control group was utilized. Although the subjects achieved a median gain of 48% in addition and 75% in division, it is impossible to say whether the improvement was due to praise, knowledge of results, or the period of practice.

Gilchrist (1916) and
Gates & Rissland (1923)

Although they utilized college students as subjects and are therefore not pertinent to this review, Gilchrist (1916) and Gates & Rissland (1923) must be mentioned because their work provided the methodology for the later Hurlock studies which formed the beginning of the great surge in the literature of interest in the effectiveness of praise and blame as incentives for school children.

Hurlock (1924)

Dividing 408 third-, fifth-, and eighth-grade children into groups by age, sex, race, and intelligence, and using praise, reproof, and control groups, Hurlock (1924) studied the effects of verbal incentives on the National Group Intelligence Tests, Scale B, Forms 1 and 2, for the fifth- and eighth-grade children and Forms A and B of the Otis Intelligence Scale, Primary Examination, for the third-graders, on a test-retest design, a one-week span between testing.

Her over-all conclusion was that neither praise or reproof was superior to the other as an incentive and that both tended to result in greater improvement than did practice.

Further analysis of the different groups suggested that older children responded more to both praise and reproof than

did the younger ones; that girls improved more from practice than did boys and were less influenced by either praise or reproof than were boys; that Negro children reacted more favorably to praise, white children to reproof. This latter difference, however, was only slight, and, according to Hurlock, showed that the two races were more alike than different in respect to their reactions to these incentives.

She felt that some incentive was more essential for superior children, if their work was to be kept up to the maximum of their ability, than for inferior ones. Both praise and reproof seemed to serve this purpose. The superior children were greatly influenced by both incentives, the inferior children decidedly less so.

Reproof has its greatest effect on the children rated by their teachers as superior in school work; praise was of greater value for children rated as average and inferior.

Hurlock (1925a)

Concerning the then current controversy over the constancy of the IQ, Hurlock (1925a), using 273 fifth- and eighth-graders, followed-up her previous work on intelligence test performance. Again she found no difference between the effects of praise and blame, but both incentives raised the IQ significantly more than did practice alone.

Hurlock (1925b)

The most distinguished work by Hurlock (1925b) on the subject of praise and blame was an attempt to study these variables over a longer period of time. "In a classroom, do the children who constantly receive praise for their work show more improvement from day to day than do the children who are reproved or who are completely ignored" (p. 146)? She used five modifications of the Courtis Reasearch Test in Arithmetic, addition form, over a period of five days, 106 children from grades four and six, the incentives administered in groups before the other members of the class, except for the isolated control group, and the following procedure: at the beginning of each

day the children in the praise group were called by name to the front of the class where they were praised for their work of the day before and urged to do even better; the reproved group was called out and severely reproved for its performance the previous day; the ignored group was present in the classroom for both of the preceding conditions; the control group was separated from the other groups after the first test.

Considering the main effects of this study, increased accuracy was found only in subjects who were praised or reproved. The greatest improvement in performance was made by the praised group, 8.41 points; the next greatest by the reproved group, 2.34 points; the ignored group improved their performance by only .54 points; while the control group showed a mean loss of .46 points. It seems possible that the control group may have experienced some difficulty in comprehending being asked to leave the room.

Side effects of this study revealed that girls showed the greatest improvement in performance under the praise incentive, boys under reproof; that there were no differences for the age variable; and that differences for the intelligence variable were small, with the lower group responding better to praise, the superior to reproof.

"These results as a whole point conclusively to the fact that regardless of the factors of age, sex, initial ability, or accuracy, praise is decidedly the most effective. . . . Reproof, when first used, seemed to be about equal in value to praise, but with continued use its effectiveness showed a decided decline" (p. 159).

Cohen (1927)

The Hurlock (1925b) experiment was replicated by Cohen (1927), with the exception that the regular classroom teacher was used, rather than an outside examiner. Less difference between the groups was found, but the same trends were apparent.

Bird (1927)

Bird (1927) gave a clinical interpretation to an analysis of factors influencing learning on a reading readiness test, based on 100 children from four- to six-years-old. Thirty were found with habitual handicaps that interfered with learning. These 30 seemed excessively dependent upon praise and for 8 of them learning performance was delayed when praise was not administered. Four children desired "to win distinction by unusual behavior. . . . To them even admonition was more satisfying than indifference" (p. 57).

"In the last analysis the same affective disturbances handicapping these children function largely in the failure of adults to achieve to the extent of their mental ability. The over-timid, the anti-social, the chronic introvert, the braggart, the annoying egotist, the excessively sociable are all victims of habits that divert the attention from the task in hand" (p. 59).

Hurlock (1931)

The first major review of the literature in this area made no attempt to criticize or comment upon the past performance of researchers, or to point out new directions for later research to explore, but was a comprehensive review of 151 experimental studies attempting to measure the influence of incentives. Six of these reviews by Hurlock (1931), in her article on the psychology of incentives, are relevant here: Binet & Vaschide (1897), Kirby (1913), Hurlock (1924, 1925a, 1925b), and Cohen (1927).

Summary up to 1930

In looking over these studies prior to 1930, the confusion of results of research studying the effects of verbal incentives on children is already apparent. In the two earlier studies, Binet & Vaschide (1897) and Kirby (1913), verbal encouragement was found to improve the scores of all participating children on both the motor task of Binet & Vaschide and the

regular classroom task of Kirby. Neither study employed a control group or attempted to compare effects of praise versus reproof.

A very careful researcher, Hurlock, in two studies (1924, 1925a), one more or less a duplication of the other, found neither praise or reproof to be superior, although both were more effective than practice. In a third study (1925b), she found praise to be decidedly more effective than reproof, although both were still superior to practice. This latter study was substantiated by a duplicate study, Cohen (1927), resulting in the same conclusions. Hurlock's first two studies used an intelligence test as the task, while the latter study made use of a regular classroom task. Bird (1927) found on a reading readiness test administered to four- to six-year-olds that for some children withheld praise caused delayed learning and that some children needed reproof to encourage their best efforts.

None of these studies with differing results used the same age levels: Kirby used third- and fourth-grades; Hurlock's two earlier studies used grades three, five, and eight, while her later study and that of Cohen used grades four and six; Bird used nursery and kindergarten children. Yet in the studies using multiple grades, no age differences were found. Could differences in methodology be responsible for such clashing differences in results?

1930-1940

Warden & Cohen (1931)

Thirty-eight fourth-grade students were subjects for the Warden & Cohen (1931) experiment on the effectiveness of schoolroom incantives. Utilizing an arithmetic problem list so long as to be unfinishable, the incentives under study--a continued story, the promise of a game, the promise of a party, verbal praise, and verbal reproof--were administered for one week each, except for praise and reproof which were given in

the same week for half a week each. For the first week the problem lists were administered without any type of incentive, with emphasis placed on accuracy rather than speed.

The speed score under the non-incentive was 16.7, the accuracy score, 8.7. For praise, the speed score was 16.6, accuracy score, 10.7. The speed and accuracy scores for reproof were 15.5 and 9.8. The best incentive was promise of a game, with speed and accuracy scores of 18.0 and 9.5. The conclusion seemed warranted by the authors, that the commonly-used incentives of verbal praise and reproof are not as effective as might be supposed for this type of routine activity in a classroom, that any change in the regular routine works as an incentive.

These results, demonstrating no difference between praise and reproof as incentives, agree thus far with Hurlock (1924, 1925a), but disagreement is obvious when praise and reproof are compared to the control group. Hurlock, using grades three, five, and eight, found both to be superior to the control group; Warden & Cohen, studying grade four, found neither superior to the control group.

Davis & Ballard (1932)

After Hurlock's review came that of Davis & Ballard (1932), presenting studies in three areas of motivation--intellectual, emotional, and social. Eight studies were included under emotional incentives, where praise and reproof were classified, only two of which are relevant to our study (Hurlock, 1924, 1925b). From their review, Davis & Ballard concluded that praise is better than reproof, that praise satiates and is better for younger children, that boys are more influenced by reproof, girls by praise, that, generally speaking, positive incentives are better than negative incentives for all ages, grades, and levels of intelligence.

Chase (1932)

A doctoral dissertation at the Iowa Child Welfare Station,

Chase (1932), studied the relative effectiveness upon comparable groups of children of a number of different types of external incentives administered under controlled experimental conditions. The subjects, 213 children from two to eight years of age, were given three series of strength tests on a hand dynamometer at one-week intervals. In the first series none of the experimental groups was told the results of its effort.

In the second series success was indicated to all but the control group, Group I, by the ringing of a bell independently controlled by the examiner. Group II was told to try again. Group III was praised. Subjects in Group IV were praised and also rewarded with a gold star by their names.

The control group, Group I, was again ignorant of the results in the third series. The silent bell indicated failure to the Group II subjects, who were asked to repeat their performance. Group III also received the silent bell treatment plus verbal reproval. Subjects in Group IV not only received the silent bell failure signal, but also had, as punishment, the red button cut off a cardboard boy with their name on it.

Results for Series II were success-repetition, success-praise, and success-reward all significantly superior to control motivation. For Series III, failure-repetition, failure-reproof, and failure-punishment all were significantly superior to control motivation. Failure-reproof and failure-repetition were significantly better than success-praise and success-repetition. This latter finding was contaminated, however, by method and previous incentive condition under the same examiner.

Chase concluded that some motivation, whether positive or negative, is more effective than none, in agreement with the two earlier Hurlock studies (1924, 1925a). The indecisive finding of failure being more effective than praise is in direct disagreement with Hurlock (1925b).

Anderson & Smith (1933)

Chase's study was duplicated by Anderson & Smith (1933) in a retest of 102 of the original 213 children. The follow-up

study, done three years later, concluded, in agreement with the Chase findings, that: (1) some motivation is more effective than control motivation, the difference significant for success-reward, failure-reproof, and failure-punishment; and (2) performance of the follow-up subjects was significantly higher when reproved for failure than when praised for success.

Additional conclusions of Anderson & Smith show significant differences between the follow-up and original studies as follows: (1) failure-repetition, which in the original study was significantly more effective than success-repetition, only approached significance in the follow-up study; and (2) failure-punishment in the follow-up study increased the original tendency toward significance to a significant difference over success-reward.

These differences between the two studies might possibly be explained by the difference in age distribution between the two, since all of the original subjects could not be utilized in the follow-up study, and those used were older than they had been previously. Necessarily there were no younger subjects in the follow-up study, and older subjects were added.

Brenner (1934)

The review by Brenner (1934), done as an introduction to his doctoral dissertation and particularly strong on philosophical positions, but also including three empirical studies relevant here (Kirby, 1913; Hurlock, 1924, 1925b), led him to conclude that the obvious conflict of results was bound to occur due to the varying experimental conditions and general inadequacies of the studies, but also "in a larger measure to the faulty formulation of the problem" (p. 48). He felt that all the studies in the literature to date erred in assuming an absolute level of difference between incentives, which "act in specific situations, depending upon all the factors of the situation as a whole . . . and derive their attributes, so to speak, from the situation in which they are active" (pp. 48-49). Brenner stated that no general conclusion could be drawn

regarding the effectiveness of praise and blame, but that their interaction with many other conditions needed to be considered. "Proper application of the incentive in a given situation depends upon the insight of the teacher. The effectiveness or worth of a teacher depends upon his ability to make adequate use of motivation" (p. 50).

Brenner himself investigated the effects of immediate versus delayed praise and blame, using a control group for both conditions and intact elementary school classes as groups--403 pupils in 12 third-grade classes, divided into immediate praise, blame, and control, and delayed praise, blame, and control, the groups matched by age, IQ, and initial spelling tests--with a spelling and recall test as the task, learning 80 words from the Ayres Measuring Ability in Spelling. He made use of repeated measures to determine if over a period of time there was some decay evident in the effectiveness of the incentive.

Although there were no significant differences between the groups, Brenner found a trend in all of the incentives toward improving the performances of second and third trials over that of the first, but after the third trial performance declined below the first trial level.

Benton (1936)

The influence of incentives on intelligence test scores of children was studied by Benton (1936), using seventh- and eighth-grade children divided into two groups of 25 each, matched by age, score on intelligence test, sex, and grade, and using equivalent forms of a group intelligence test administered 28 days apart. The experimental groups were given praise, promise of prizes, and information on standing, but did not differ significantly from the control group which had none of these conditions.

Forlano & Axelrod (1937)

Forlano & Axelrod (1937) conducted the first study which attempted to take into account personality differences. At the fifth-grade level, with praise, blame, and control groups,

they divided 131 subjects into introvert and extrovert groups on the basis of the Pintner Inventory, and used the Woodworth-Wells Number Cancellation Test as the task. After the first test was administered, the students in the praise and blame groups were called individually to the teacher's desk where their papers were marked P or G, not according to performance, but by a pre-arranged schedule, the pupils being instructed not to show their papers to others. The examiner then told the class as a whole that those whose papers were marked P did poor work and those who received a G had done good work, on the theory that in this manner each child received individual praise or blame from the teacher and not group reward.

A second form of the test was administered for two minutes, after which the subjects received the same incentive mark as they had received originally. A third form was administered and collected without any marks or remarks. The control group was administered the tests without the incentive marks.

All groups showed significant gains on the second and third trials, with the blame groups, both introverted and extroverted, having the greatest mean differences, although the variability was also greatest in both cases. Only two differences were significant for the second trial: between introverted blame and introverted praise and between introverted blame and introverted control. The authors concluded blame to be a more effective incentive than praise, but speculated on the increase of the standard ratios for the praise groups and whether more trials would have brought about significance.

Murphy, Murphy, & Newcomb (1937)

Murphy, Murphy, & Newcomb (1937), in their chapter on aggression and competition, list and digest in a series of tables a large number of studies concerned with incentives, eight of which are relevant here (Kirby, 1913; Hurlock, 1924, 1925b; Warden & Cohen, 1931; Chase, 1932; Anderson & Smith, 1933; Brenner, 1934; Benton, 1936). Forlano & Axelrod (1937) they reviewed in their text. Their chief concern was the problem "of the relationship between the individual personality and

the situation. . . . All of the experimental work attempting to analyze the effects of praise and reproof . . . was until recently carried out without any references to the values or meanings of the situation to the individual child and its relation to his own personality" (pp. 449-450). They call attention to the inconsistencies in the literature and attribute this confusion to a failure to consider the individual in relation to the cultural background and his own peculiar personality make-up.

Schmidt (1941)

An excellent review of the literature by author was done by Schmidt (1941), although he discussed only a few of the studies, among which were four previous reviews (Hurlock, 1931; Davis & Ballard, 1932; Brenner, 1934; Murphy, Murphy, & Newcomb, 1937) and three pertinent studies (Binet & Vaschide, 1897; Bird, 1927; Forlano & Axelrod, 1937). From his review Schmidt drew the following conclusions:

(1) "The effects of praise being a greater incentive than blame, or vice versa, is still an unsettled question.

(2) "The later studies tend to show little difference between the effectiveness of these incentives, and if leaning in any direction, do so towards blame.

(3) "Both praise and blame give increased performance on the second and third trials. Additional trials will produce one of two results, dependent upon the study chosen: either a continuing increment or a decrement in performance.

(4) "It appears that both praise and blame show only slightly increased performances in the second and third trials over any other change that might be introduced into the learning situation.

(5) "Really adequate comparisons cannot be made between studies because:

The variables of IQ, CA, MA, grade, initial ability, intensity of motivation (or type), and Emotional Aspect, have not been equated for, or taken account of, in many studies, or are on entirely differing planes.

Techniques, approaches to the question, test media, vary considerably between experiments.

Often the statistical treatment of the data has been unreliable or inadequate.

Generally, a lack of any common ground upon which to made comparisons.

(6) "The actual motivating factor has been accepted too casually (with the possible single exception of Brenner):

Equal statements of praise or reproof are not necessarily of equal weight in effectiveness; what might act as blame for one individual (donor or recipient) might act as praise for another.

The method of administering the incentive will affect its potency, and has varied considerably from experiment to experiment; from a simple statement of praise or reproof to a quite lengthy statement, and from the giving of it to the individual alone to the giving of it to the group or to an individual before a group.

(7) "The <u>total</u> situation, including the physiological, has been given little attention.

(8) "In none of the studies has the factor of the influence of the experimenter's personality been taken into account.

(9) "And, finally, there has been no distinction made between learning and the ends employed to seek its measurement.

"This review of the literature leaves no definite conclusions to be drawn as to the effectiveness of praise and blame as incentives to learning. This has been brought about, it is felt, through there being no clear-cut conception of the basic principles involved, or through a non-realization or non-differentiation of what was being sought and what was actually being done. It is true that performance has been mentioned frequently in the literature cited, but this has been done casually and interchangeably with learning" (pp. 6-7).

Summary 1930 to 1940

On one point all of the studies in this decade agree: the experimental incentives were administered regardless of the actual performance of the children. This in itself is a somewhat artificial approach, since classroom teachers would supposedly be using the incentives to fit the occasion. But whether or not the answers were correct or incorrect, or the performance satisfactory or unsatisfactory, during this decade, as for the previous pertinent research, subjects were administered the incentives at random. Warden & Cohen (1931) not only found praise and reproof to be not significantly different from each other or the control group, but also to be inferior to promise of a game in motivating fourth-graders in a regular classroom task. Brenner (1934) also found no differences between praise, blame, and control groups, even when divided into delayed and immediate praise and blame. Nor did Benton (1936) find any differences between praise and control: he did not make use of a reproved group.

Yet Chase (1931), using a motor task, found both praise and reproof to be better than the control, and failure to be better even than praise. However, this latter finding was contaminated by previous incentive condition and method of administering the task. Anderson & Smith (1933), in a follow-up study, agreed with Chase's finding and definitely concluded blame to be more effective than praise. And Forlano & Axelrod (1937), endeavoring to take into consideration the personality of the subjects, found a significant difference only in the introverted subjects, where blame was more effective than praise or control. There were no differences in effectiveness of incentives for the extroverts.

None of these later studies agreed with Hurlock (1925b) that praise is more effective than reproof. The tendency for this period of research leans either to no differences between praise and reproof, or reproof to be more effective. There is still no generally accepted conclusion from most researchers in the field of effectiveness of verbal incentives for school

children. It is no wonder that researchers are beginning to end their reports with statements to the effect that praise and blame must be used wisely and with understanding for the specific pupil at the specific time.

1940-1950

Potter (1943)

The effect of reproof at different age levels was studied by Potter (1943), using 234 school children in grades three, six, nine, and twelve, half of each class as experimental, half as control, and applying the reproof in connection with an arithmetic task repeated three times to both groups, the control group told to retake the test each time, the experimental group verbally reproved on their performance before the second trial and then told to retake the test, and on the third trial simply told to retake the test.

Both groups improved their scores. Intergroup comparisons revealed that, with the exception of grade three, the experimental group showed greater improvement. In grade three reproof impaired performance, improved it in grade six, and had slight effect in grades nine and twelve. Potter concluded that at higher ages reproof had little effect on performance, while at lower age levels the effect varied.

Klugman (1944)

Making use of the equivalent forms provided by the 1937 revision of the Stanford-Binet, Klugman (1944), using a test-retest procedure, studied the effects of praise and money as incentives on 72 children from grades two to seven, the exact age distribution unstated. Klugman concluded that there was no significant reliability difference between the reward conditions; that there was a substantial superiority of the scores of Negro children tested with the money incentive over those tested with the praise incentive; and practically no difference between Negro and white children when money was the incentive; that which children showed a substantial superiority over Negro

children when praise was the incentive. He accounted for the lack of significant differences between the groups on the basis that the task was "so challenging that the use of one incentive rather than the other made little or no significant difference" (p. 268). When evaluating Klugman's results it must be taken into account that he did not make use of a control group.

Thompson & Hunnicutt (1944)

Thompson & Hunnicutt (1944), like Forlano & Axelrod (1937) earlier, were interested in the effects of praise and blame on the work achievements of introverts and extroverts. The methods of the two studies were similar, with the later study testing its subjects, 124 fifth-grade students, over a longer period of time, six repeated measures under the reward conditions as opposed to three trials in the earlier study. The subjects were divided into introvert and extrovert groups, through use of the Pintner IE Scale, and a control group. Six alternate forms of a number cancellation test were constructed to measure each pupil's performance over a span of six administrations of 30 seconds each.

Both of the total praise and blame groups were significantly superior to the total control group, causing the authors to conclude that the effectiveness of verbal incentives in motivating pupils was more noticeable when introverts and extroverts were not differentiated. But there was no significant difference between the praise and blame groups.

Unlike Forlano & Axelrod, who found introverted blame to be significantly superior to introverted control and introverted praise, Thompson & Hunnicutt for the first three trials found no significant differences between groups. On the final three administrations of the test, repeated praise increased the performance of the introvert praise group until it was superior to introvert blame and extrovert praise, answering in the affirmative Forlano & Axelrod's question as to whether repeated trials would raise the effects of praise to significance. Repeated blame increased the performance of the extrovert blame group until it was superior to extrovert praise and introvert blame. On

none of the tests did the extrovert praise or introvert blame groups differ significantly.

Thus, for this study of fifth-graders divided into introverted and extroverted groups, application of praise and blame, when repeated often enough, had differential effects: introverts achieved a higher level of performance when praised, extroverts when blamed. Reminiscent of Brenner (1934), the authors commented upon this conclusion by stating, "the results of this study indicate that praise, as well as blame, can be used unwisely by the elementary school teacher if he does not fully appreciate and understand the different personalities present in his classroom. Praise and blame should not be judged on an either-or basis, but should be used to fit the case" (p. 266).

Summary 1940 to 1950

Whether it was the effect of World War II, a shift in interest by investigators in the field, or high standards for the next step set by Schmidt (1941), the decade 1940-1950 added only three studies to the investigation of the effects of praise and blame on the performance of school children. Only one of these, Thompson & Hunnicutt (1944), endeavored to compare the effects of both praise and blame on student subjects.

Potter (1943), studying the effect of reproof on the performance of children at different age levels, found that it decreased performance of third-graders, increased that of sixth-graders, and had little effect on ninth- and twelfth-graders.

In an effort to establish a relationship between praise versus a money reward as incentives, Klugman (1944) found no differences in performance on the Stanford-Binet. His is the first study since that of Bird (1927) which does not employ a control group.

In their replication of the Forlano & Axelrod study (1937) on the effects of praise and blame on introverts and extroverts, Thompson & Hunnicutt (1944) discovered that adding three more trials to the original three trials changed the

significant results of the study. Introverted praise and extroverted blame were significantly superior to introverted blame and extroverted praise in the later study, whereas in the original study only introverted blame was superior to introverted praise. Forlano & Axelrod could only conclude that blame was better for introverts. On the basis of repeated praise and blame Thompson & Hunnicutt concluded praise to be better for introverts and blame to be better for extroverts.

1950-1960

Mech, Kapos, Hurst, & Auble (1954)

Mech, Kapos, Hurst, & Auble (1954) did the first study in the 1950-1960 decade to relate specifically to the problem at hand, although there began to appear in the literature a parallel line of investigations using knowledge of results as reinforcement, rather than praise and blame. There is, of course, a certain amount of overlap, but for reasons of clarity, simplicity, and emphasis, this series of studies will not be discussed.

Mech *et al.*, to test whether praise was more effective when used in massed or spaced distributions, studied 240 fourth-grade subjects divided into massed versus spaced reinforcement groups, further divided into 0%, 50%, and 100% verbal reinforcement. The task was 14 trials of 4 minutes each on an arithmetic problem list so long as to be unfinishable, the verbal reinforcement, praise, given in front of the class.

The 14 trials were given one after another in the massed reinforcement group, and at intervals of 10 minutes in the spaced group. Results of an analysis of covariance revealed that the massed-trained group performed significantly better. It was concluded that the inter-trial interval is a factor determining the effectiveness of praise.

It is unfortunate that in this study, reward condition is confounded with percentage of reinforcement. The control group of 0% reinforcement would have seemed to be an important

factor here in determining any effect of praise at all. Also there is the question of a fatigue factor. One group finished in approximately an hour, the other in three hours.

Kent (1956)

In his study to determine if the motivational needs of blind children are similar to those of sighted children and if for blind children they are similar for residential-school blind as for public-school blind children, Kent (1957) used 60 residential-school blind children, 43 public-school blind children, and 60 public-school sighted children in experimental groups, plus 30 residential blind as a control group. All subjects were classified as introvert or extrovert by means of the Pintner Scale. An adapted version of the Woodworth-Wells Number Cancellation Test was administered six times with alternate forms. The score difference between the first and sixth trial was used as the measure of effectiveness.

Kent hypothesized, from results of the Thompson & Hunnicutt study (1944), that introverts would respond more readily to praise, extroverts to reproof, but found that either incentive was better than none for all groups except the public-school blind, where there was no difference between incentives and control. Praise was found to be more effective regardless of personality type or visual status.

The degree of achievement between residential blind and sighted public school children was not significantly different, but both of these groups significantly exceeded the public-school blind.

Silverman (1957)

Studying the effect of praise and reproof on reading improvement in a regular classroom setting, Silverman (1957) used 904 school children from third-, fourth-, fifth-, and sixth-grade classes: 23 third, 13 fourth, 9 fifth, 4 sixth. Forty-nine first-year teachers were selected from 1628 student teachers of 1953-54. Six observers visited all 49 teachers at beginning and end of experiment, one observer in the classroom at a

time, the teachers expecting visitors but not knowing what specifically was being observed.

Using Olsen's time-sample technique, dividing the half hour into six five-minute observation periods, the observers tallied each remark the teacher made in one of five categories: (1) supportive statements, (2) problem-structure statements, (3) neutral statements, (4) corrective statements, and (5) reproving statements. During the first visit the observer made a new tally mark only when the teacher shifted from one category to another. During the second visit the frequency of occurrenc of statements in each category was tallied. One tally was made each five minutes for the appropriate subject matter area in which the class was working.

Reading ability was measured at beginning and end of experiment by the California Reading Test. Form S of the California Test of Mental Maturity, non-language section, was administered at the beginning of the school year. Rate of improvement on the reading test was a dependent variable adjusted by covariance for intelligence and initial reading scores.

It was concluded that teacher use of praise and reproof cannot predict the reading growth of pupils, but can predict total verbal output. Silverman postulated atypical teacher behavior due to presence of the observer. One methodological disadvantage of this study, along with those of Potter (1943) and Brenner (1934), was the use of intact classrooms and thus the inability to randomly assign subjects.

Terrell & Kennedy (1957)

A variety of reward conditions (praise, reproof, candy, delayed reward, and control) were studied by Terrell & Kennedy (1957), using 160 subjects at two age levels, four- and five-year-olds and eight- and nine-year-olds, on a simple discrimination task. The dependent measure was the number of trials to a criterion of nine correct responses on the last ten trials.

All subjects were rewarded by a light flash after each correct response. The praised group was verbally praised after

each correct response. Reproof was given after each incorrect response. The candy reward was given after each correct response, and the delayed reward group, after each correct response, was given a dried bean to accumulate and later exchange for a bag of candy. After the criterion was reached, the subjects were given a four-trial transposition test under the same incentives.

An analysis of variance, run on the number of trials to criterion and the number of correct responses on the transposition test, indicated that the candy group had learned the task significantly more quickly than the other groups and had transposed this learning significantly better than had the other groups. Neither praise or reproof was significantly more effective than the control group. There were no differences between age groups on transposition, however the older children learned significantly more quickly than did the younger.

Ellis & Distephano (1958)

Ellis & Distephano, investigating the effect of verbal praise on the acquisition of rotor pursuit skill in mental defectives, matched 28 children in experimental and control groups by Binet IQ, age, sex, and previous pursuit rotor performance. The significant difference between the groups led the authors to conclude that from verbal encouragement and praise there is a definite facilitating effect upon the psychomotor performance of mental defectives.

Sandstrom & Weinz (1958)

Using 83 twelve-year-old boys as subjects, Sandstrom & Weinz (1958) studied the effect of praise and reproof on a localization problem--four responses with each hand, using a mapping pin to mark the underside of a table to correspond to a black fixation point on the table's surface. After the initial trial no information was given regarding accuracy of the performance. One week later a second trial was administered under the incentive conditions. The reproved group was told it had done badly the first time. The praised group was not only told that

it had done well the first time, yet could do better, but that for each direct hit a crown would be awarded.

Under these conditions the reproved group's second response was more stable than its first, whereas the praise group did more poorly on the second trial, yet was still significantly different from the control group. The authors concluded subjectively that the praised group was inferior to the reproved group because of the conflict between encouragement and the unattainable reward.

Summary 1950 to 1960

The decade 1950-1960 doubled the number of studies of the previous decade, and introduced interesting variations to the research theme. Blind children and mental defectives were studied (Kent, 1956; Ellis & Distephano, 1958). The experimental incentives were administered according to the performance of the subjects (Terrell & Kennedy, 1958). A localization task was used (Sandstrom & Weinz, 1958), and massed versus spaced praise was studied (Mech, Kapos, Hurst, & Auble, 1954). Teacher use of praise or reproof in the classroom was recorded in an effort to evaluate its effect on reading improvement (Silverman, 1957).

Results were just as varied, but almost predictable: massed praise was found to be more effective than spaced praise; even for blind children praise and blame were more effective than control, except for public school blind where there was no difference between groups and where performance was itself significantly inferior to that of residential-school blind and public-school sighted children; teacher use of praise and reproof in a classroom cannot predict reading growth in pupils; candy is a significantly more effective reward than praise or reproof, neither of which is better than a knowledge-of-results control group; praise is significantly more effective for mental defectives than no incentive; reproof is better than praise if the reward offered is unattainable, although both are better than no information on results of performance.

Although one study found reproof to be more effective,

the conclusion was qualified by experimental conditions in the praised group. No really radical findings are noted for this decade. Possibly its importance lies in the varied forms that the research was taking, in an evergrowing effort to try to reach some valid decision as to the effects of praise and reproof on the performance of school children.

1960 to Date

Dollins, Angelino, & Mech (1960)

To study the effect of teacher praise upon 75 elementary school children below the 30th percentile on the California Test of Personality (Elementary Form AA), Dollins, Angelino, & Mech (1960) used three experimental groups and arithmetic problems administered for three minutes each day for 30 days. One group received 20 days of praise out of the 30 testing days, one group was praised 10 out of the 30 days, and the control group received no praise at all. After the California Test of Personality was re-administered, no significant differences were found between the three groups, nor was there any difference in the arithmetic scores of the three groups.

Stevenson & Snyder (1960)

Stevenson & Snyder (1960) investigated the possible interaction effects of three types of incentive conditions on the performance of 90 familially mentally-retarded children who had a mean CA of 13.5 and a mean MA of 6.5, and who demonstrated no gross sensory motor or emotional disturbances. At first the children were divided into three groups--praise, reproof, and control--to play the first game, putting marbles in a hole in a box. Reinforcement was given every 30 seconds in the 7-minute game. The second time the game was played the original three groups were each subdivided into three groups--praise, reproof, and control.

For Game 1 the control group had the highest mean number of marbles inserted into the box; the praised group was next;

the reproved group lowest. For Game 2, the subjects who had the same incentive as for Game 1 had the same results. However, when the incentive changed the results changed.

This study provided clear evidence that the effect of a particular incentive condition is influenced significantly by the type of incentive condition that preceded it. Reproof followed by control led to depressed performance. Reproof followed by praise resulted in increased performance. When praise was the initial incentive, more praise resulted in a rise in performance, but reproof or control following praise did not lead to a decline in performance.

The authors concluded that when the examiner's statements lead subjects to assume that a particular level of performance is expected or that his performance is less satisfactory than that of other subjects, failure increases motivation; but when the examiner's statements only comment upon subjects' performance, failure lowers motivation.

Fischer (1962)

Studying the relationship between concrete and abstract reinforcement on the sharing of preschool children, Fischer (1962) used 24 nursery school children with a mean age of 50 months and a mean IQ of 126. Bubblegum was found to be a more effective incentive than praise for encouraging sharing in this aged child, a finding consistent with Terrell & Kennedy (1957), who also found an immediate food reward superior to verbal incentives.

Kennedy, Turner, & Lindner (1962)

To analyze the effects of praise and blame as a function of intelligence, Kennedy, Turner, & Lindner (1962), in a study which directly stimulated the present investigation, used a discrimination task involving an oddity in one of four patterns. Subjects were 96 eleventh- and twelfth-grade students, 48 with an IQ range from 124 to 150 from an NSF-FSU Summer Math Camp and 48 with an IQ range from 95 to 116 from a high school. Each group was randomly subdivided into three treatment groups:

praise, blame, control. For the superior group no difference in performance was found between the three incentive groups, while for the average group there was a significant difference resulting in a decrement in performance for the blamed group.

Metz (1962)

To study the relative effects of stress and praise on creativity, Metz (1962) used 108 ninth-graders performing on selected tests of the Guilford tests of creativity. All subjects were administered the high school form of Mandler and Cohen's Test Anxiety Questionnaire, the Taylor Manifest Anxiety Scale, and nine tests of creativity, as well as a test attitude questionnaire. The subjects were grouped by high and low intelligence, high, medium, and low anxiety, and three experimental conditions, stress, praise, neutral.

Stress (told they had failed the test) facilitated performance of low scholastic aptitude subjects, inhibited high scholastic aptitude subjects. Praise facilitated performance of low aptitude subjects. Neutral instructions and low aptitude were associated with poorest performance. High aptitude subjects who received either praise or neutral instructions achieved the highest performance and performed equally well under either incentive.

Tiber & Kennedy (1963)

Tiber & Kennedy (1963), studying the effects of incentives on the intelligence test performance of different social groups, used 480 second- and third-graders divided into three social groups (middle-class white, lower-class white, and lower-class Negro), each randomly subdivided into four incentive groups--praise, reproof, candy reward, and no incentive. The 1960 Stanford-Binet Form L-M was used with experimental incentives administered at the end of each subtest.

Contrary to expectations, the statistical analysis indicated no difference between incentive groups and no significant interaction between type of incentive and social group. However, as in previous research, different class and caste

groups did produce significantly different mean IQ scores. The authors suggested that explanations of IQ differences between cultural groups must be based on causes other than lack of intrinsic motivation provided by the intelligence test itself.

Willcutt & Kennedy (1963)

Ninety lower-middle and upper-lower class fourth-grade students of high, medium, and low intelligence were divided into three verbal incentive treatment groups, praise, reproof, and no incentive, by Willcutt & Kennedy (1963), who, using the same task and procedure as Kennedy, Turner, & Lindner (1962), further studied the effects of praise and blame as a function of intelligence. At this grade level there was no significant relationship between intelligence and effectiveness of verbal incentive. There was a significant difference at this age level, however, between the main effects, praise being superior to either reproof or no incentive.

Summary 1960 to date

Five of the seven studies in the early 1960s pertain directly or indirectly to the problem of establishing a relationship between verbal incentives and intelligence (Stevenson & Snyder, 1960; Kennedy, Turner, & Lindner, 1962; Metz, 1962; Tiber & Kennedy, 1963; Willcutt & Kennedy, 1963).

Stevenson & Snyder, with a more complicated research design than that of Ellis & Distephano (1958), found that for mentally-retarded children changing the incentives changed their effect on the subjects: praise as the original incentive maintained higher performance even when followed by blame or no incentive, while with more praise performance was increased; blame followed by praise led to increased performance, but a decrement in performance was apparent for blame followed by no incentive.

Metz, incidentally studying effects of incentives on intelligence, found both stress (subjects told they had failed) and praise to facilitate the creative performance of inferior

students. For superior students creative performance was inhibited by stress, while either praise or neutral instructions led them to superior performance.

Two studies in this decade illustrated a difference in the interaction between verbal incentive and intelligence by using the same discrimination task and procedure at different grade levels. For eleventh- and twelfth-graders Kennedy, Turner, & Lindner found no difference between the effects of praise, blame, or no incentive on the superior students, yet for students of average ability in the reproved group, a significant decrement in performance was obvious. Willcutt & Kennedy found that for all fourth-graders, regardless of intelligence level, praise was the most effective incentive.

Changing the task from a discrimination problem to performance on an intelligence test, the grade level to second and third, and controlling for class and caste social groups rather than level of intelligence, Tiber & Kennedy, contrary to expectations, found no difference in effectiveness of incentives for any of the social groups, yet found a significant difference in IQ scores between these social groups. The question arises: was it the difference in task, the difference in grade level, or the confounding of intelligence with social groupings that was responsible for these results?

At any rate, these studies seem to suggest that in some way there is an interaction between intelligence, age, even social group, and the effects of praise and blame on school children.

The study of Dollins, Angelino, & Mech (1960), which concluded teacher use of praise in the elementary classroom did not effect student personality or performance on an arithmetic task, also raises an interesting question. Could it be that the finding of Mech, Kapos, Hurst, & Auble (1954), that massed praise is more effective than spaced praise, has a bearing on the discrepancy between the Dollins-Angelino-Mech study and the Willcutt-Kennedy study? The first study was spaced over 30 days with praise at its maximum given on only 20 of those days, while

the Willcutt-Kennedy study, for the individual child, was completed in only a short time on one day, with the praise being given immediately for a specific task at hand.

The results of the seventh study in the early 1960s are in agreement with earlier research: Fischer (1962) found, just as did Terrell & Kennedy (1957), that an immediate food reward is the most effective incentive for youngsters.

Over-all Summary

In looking back over these studies the most apparent problem is that although results and conclusions are different and diversified, so are problems, and methods, and subjects, and tasks, and method of statistical analysis, and factors controlled for and factors not considered, and so on almost ad infinitum. Even after 65 years and 29 studies endeavoring to shed light on the relative effectiveness of incentives for school children, it can not be said with authority: "This is so, for this particular aged child at this particular level of intelligence in this particular social group." The definitive study has not yet been done.

Problem

Previous research by the writer (Kennedy, Turner, & Lindner, 1962; Willcutt & Kennedy, 1963) indicated praise to be more generally effective than blame as a verbal incentive for children, but that the effectiveness varied with grade level and intelligence. Although the study by Tiber & Kennedy (1963) did not support the hypothesis of an interaction between race and incentive that earlier studies had suggested (Hurlock, 1924; Klugman, 1944), it did add to the list of studies which found a significant difference between the intelligence test performance of different racial and social groups. Following this work and a survey of the literature it was felt that there were several factors which contributed to the varying effectiveness of praise and blame: grade, sex, intelligence, race, examiner, school, and social class.

Because of the very nature of the classroom setting there is little question that verbal praise and blame are the most often used incentives in the armamentaria of the teacher. These incentives have a history of long use, as well as a large literature on both a philosophical and empirical basis. Often, however, the incentives have been considered one at a time, and even in studies where both were considered, the comparability of findings is open to question due to the lack of uniformity in methodology and conditions. What seems to be needed now is a study which attempts to take into account those variables which are known or have been suggested to have a bearing on the effectiveness of verbal incentives with school children.

The present study was an investigation of the effects of praise and blame on a discrimination task under the variables of grade, intelligence, sex, race, social class, school, examiner. Discrimination and learning were selected as the dependent variable because of the relevance to a classroom setting. All of the subjects could solve the task used in this study, but the speed with which it was solved seemed highly variable. The analysis of this variance was the major statistic of the study.

Method

Instruments

Four instruments were used in the investigation: the California Test of Mental Maturity, a Wollensak tape recorder, an Esterline Angus 20-pen event recorder, and a four-key discrimination box designed for the study.

The sample was divided into three levels of intelligence through the administration of the California Test of Mental Maturity. It was felt that intelligence might effect the criterion measure or other variables in the study. The Short Form A was used because of the difficulty in scheduling the full test during classroom time. This particular test was used because of its wide use in the school system and because it would provide the gross classification needed: only the total IQ was used for this study.

The Wollensak tape recorder was a standard magnetic, two-speed model T1500, 115 volts, 60 cycles, with a standard crystal microphone. The entire procedure was monitored: both the examiner's instructions and the verbal response of the subject.

The Esterline Angus 20-pen event recorder was a voltage type, separate circuit, 120 volt drive, and pen rating. The chart drive was 12 inches per minute, the chart calibrated in tenths of seconds. This recorder was connected directly to the four-key discrimination apparatus and monitored the examiner's and subject's responses by recording from the four subject keys and the single examiner key.

Constructed of 16-guage aluminum, the four-key discrimination apparatus is 20 inches long, 6½ inches wide, and 8 inches tall. The 9 x 6½ inch front panel slopes backward 4 inches at the top, and has a viewing screen 4 inches wide and 2½ inches tall, centered 1 inch from the top. Four double-pole, single-throw, momentary keys are 1½ inches below the screen. A 5 x 5 inch card slot parallel to the viewing screen is 7½ inches behind the screen. Between the screen and card is a polarized glass at a 30° angle to prevent the card from being seen until a light below the glass is turned on.

The experimenter's panel is concealed from the subject by a cardboard screen. On the panel are indicator lights for each of the four subject key switches. Also mounted in the panel is a standard chronoscope calibrated in hundreths of a second, and a single-pole, single-throw, momentary switch which simultaneously starts the clock and turns on the view screen light. The circuit is wired, with a 5 amp fuse on both sides, so that the clock and light continue to operate until the subject depressed a key, which de-activates the clock and light and indicates on the control panel which key was depressed.

Illustrated in Figures 1 and 2 are the testing complex and its wiring.

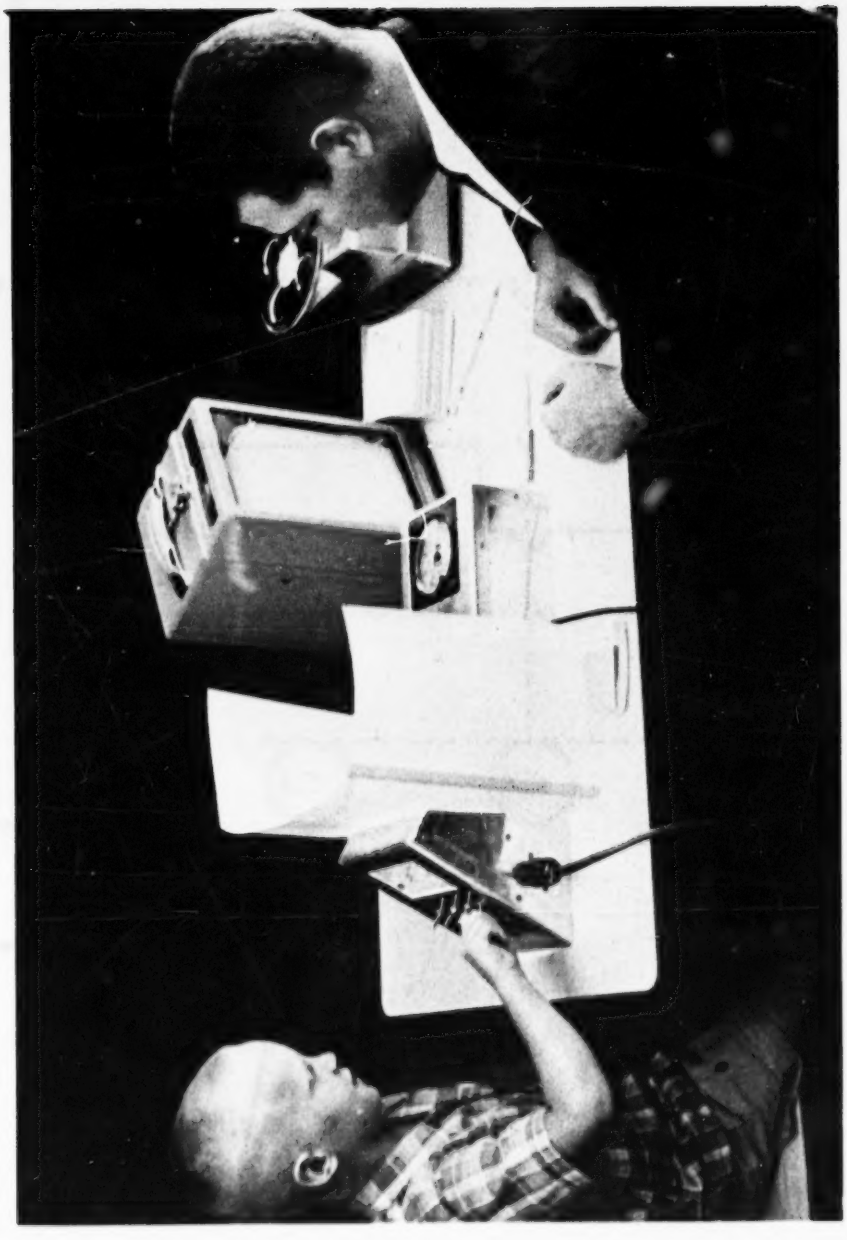

Fig. 1.--Testing Complex

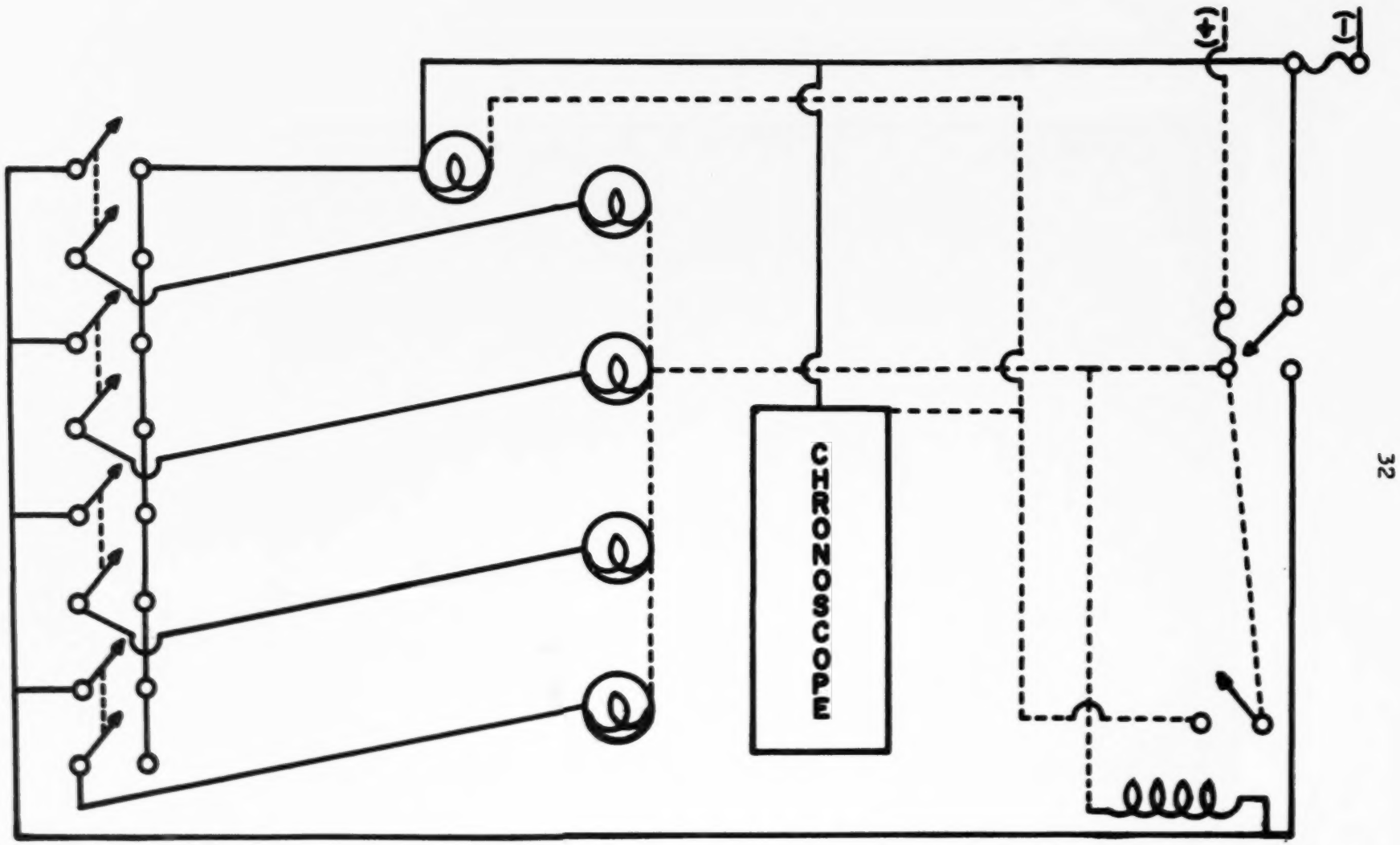

Fig. 2.--Diagram of Apparatus Wiring

Stimulus Cards

The 32 oddity-problem stimulus cards, identical to those used by Kennedy, Turner, & Lindner (1962) and Willcutt & Kennedy (1963), presented four patterns, one of which was different from the other three. The patterns were created as follows: one-inch squares were divided into 64 squares, 32 white, 32 black, the location of the black and white squares making the different patterns. Each original one-inch square was divided into four quadrants of 16 squares each, 8 of which had to be black. Each quadrant was further divided into sub-quadrants of 4 squares each, 2 of which would be black. The placement of these 2 black squares was done randomly, because pilot work had indicated that random assignment for a complete quadrant resulted in non-symetrical patterns which significantly effected the subject's ability to discriminate. That is, all too often the black squares were adjacent to each other, so that change in configuration was too easily located.

The fourth pattern, or odd pattern, for each stimulus card was formed by specific modifications involving eight levels of discrimination difficulty established first on a theoretical basis and then empirically (Turner, 1962). The most difficult discrimination was formed by moving only 1 of the 64 squares from one quadrant to another. The square to be moved and quadrant to which it was moved were selected by a table of random numbers. The second most difficult problem involved moving 2 squares, and so on down to the easiest problem of 8 squares being moved from one quadrant to another. On these levels of difficulty a restriction on random placement was upheld in order to keep a balance between the remaining three quadrants. Thus, when 3 squares were being moved, 1 was put into each of the three remaining quadrants.

In each case the odd design was randomly assigned to one of the four positions on the stimulus card, four cards being constructed at each of the eight levels of difficulty. The 32 stimulus cards are illustrated in the Appendix.

Initial Procedure

Four identical sets of apparatus, each used at a different grade level, were constructed and sent into the field. Data was simultaneously gathered for grades two, four, seven, and ten. All tests were administered in private in a quiet room at the school of each subject. Table and apparatus height was adjusted appropriately for each subject. After initial data gathering--name, age, grade, and classroom teacher checked against prepared roster--two cards of the Thematic Apperception Test were administered. This procedure, the results of which are described elsewhere (Bluhm, 1963), was introduced to determine if there was a relationship between anxiety level and reaction to level of difficulty, and took approximately five minutes. Then all subjects were read the following directions:

> This game is to see if you can tell the difference between designs. When I turn on the light, you will see four designs there in front of you. (Examiner shows subject where to look.) You are to find the one that is not the same as the others. Three of the designs will be the same, but one will be different. As soon as you find the one that is different, press the handle underneath it. For example, if the first design on the left is different from the rest, you press the first handle on the left. If the second design is different, press the second handle. And so on. Have you any questions?

The examiner then demonstrated a sample card to make sure that the subject understood the design-lever combination. Although the difficulty of instructions was perhaps too high for second-graders, the combination of instructions and demonstration was effective: the error variable was virtually non-existent. With occasional and apparently random exceptions, the subjects in the experiment correctly discriminated all of the problems. Following the demonstration, instructions were continued:

> Now you try the handles yourself. Press number one, please, two, three, and four. O.K. The first card is just for practice so you can see what the designs look like. You press the handle under the design that is different from all the rest. Ready? (If the subjects pressed the correct lever, the examiner said, "good.")
>
> All the cards will not be the same, but the problem will be the same. That is, to find the one design that

is different from all the rest on the card. Any questions?

Put both of your hands palm down on the table in front of the handles. Keep them on the table until you are ready to press the handle. Do not jump ahead. Press the handle as quickly as you can, but be sure to press the one under the design that is different. Ready?

If the subject had further questions, the appropriate portion of the directions was reread. If necessary, a second sample card was used. The examiner was as active as necessary to assure that the subject fully understood what was required of him.

Trials

A trial consisted of 32 stimulus cards ramdomly ordered at each difficulty level, but in sequence as far as difficulty was concerned. That is, the easier discrimination problems were presented first, building up to the more difficult. Or more specifically, the subjects started at difficulty level one, worked up to level eight, and then back to level one, to repeat the pattern through the 32 cards.

At the end of trial one reproof was administered to subjects assigned to the reproof group as follows:

Well, that's it. (Pause) Your scores are not nearly as good as I thought they would be. I'm really disappointed. I think these are the worse scores I've gotten yet. (Pause) But perhaps if we ran through the cards again. I'll tell you what. Just forget these scores. Let's go through the cards again. Do you understand the directions? Let me repeat them once more.

The directions were then reread verbatim and the same 32 cards re-administered in the same order.

Subjects in the praised group were told:

Well, that's it. (Pause) Your scores are very good. I knew you would do well, but you did even better than I expected. I am really pleased. These look like the best scores I have gotten yet. (Pause) Now, I'd like you to go through the cards again. To get the time the same, I'll repeat the directions.

Subjects in the control group were asked questions of a biographical nature during the interval between trials. The time to do this was equated as closely as possible with that used to present the verbal incentives.

At the end of the second trial all subjects were asked not to discuss the experiment with anyone. Those who served in the reproof group were detained and given high praise for their two performances to counteract any adverse effect resulting from the reproof.

The examiner assignment was such that each examiner did approximately the same number of subjects from each of the reward conditions. However, the interaction between examiner and reward condition was analyzed.

Criterion Measure

The reaction time between the stimulus cards appearing on the screen and the depression of the subject key was the criterion measure for this experiment. The major variance studied was the mean reaction time (MRT) of trials one and two and the MRT difference between trials one and two.

Statistical Design

Lindquist repeated measures analysis of variance was used to analyze data in terms of reward condition, grade level, level of intelligence, and sex. Because of the difficulty in equating for intelligence, Negro and white data were analyzed separately. There were three reward conditions: praise, reproof, control; four grade levels: two, four, seven, ten; and three levels of intelligence: high, medium, low. There were 5 subjects in each of 144 cells.

Illustrated in Figure 3 is the analysis of variance cell division.

In addition to these variables, the data were also studied in relation to socio-economic level, examiner variability, and variability between schools used in the study. An attempt was made here to assess the stability of the measure.

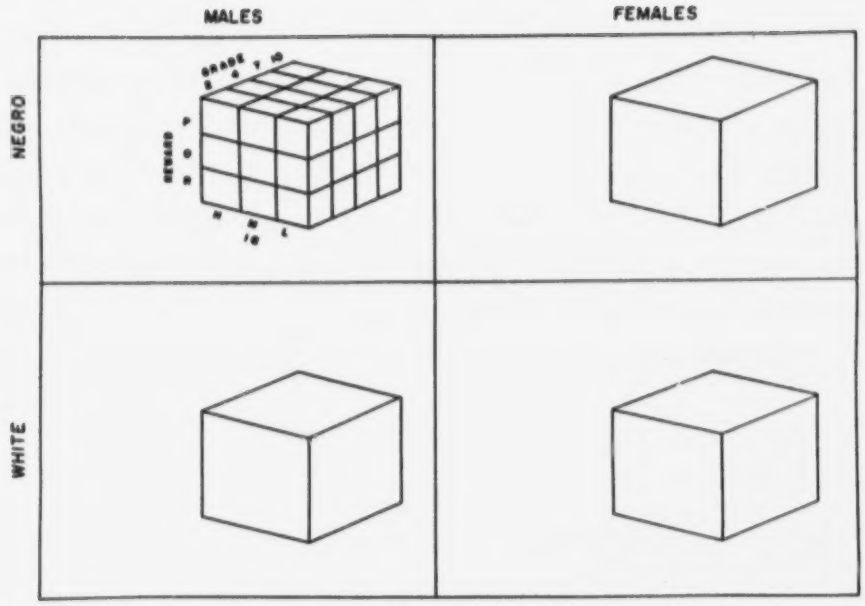

Fig. 3.--Analysis of Variance Cell Division

General Population Characteristics

The sample for this study was drawn from a large southern metropolitan area (population 314,000), which has shown a 36% population increase between 1950 and 1960. Thirty-two per cent of the population is non-white, mostly Negro, and 41% of the population is under 18 years of age. Using the 1960 occupational distribution schedule from the U.S. Census, divided into seven classifications according to the McGuire-White Scale, a comparison between the major occupational groups of the parents of the children in the sample and those of the parent population is presented in Table 1. The McGuire-White Scale rating of 1, the highest occupational classification, includes lawyers, judges, top executives, newspaper and magazine editors, etc. On the lowest level, 7, are the heavy laborer, odd-job man, miner, unskilled worker, domestic help, etc.

Table 1

Distribution of Occupations for Parent Population and Families of Children in the Sample

McGuire-White Scale Value	Parent Population Percentage		Sample Population Percentage	
	Negro	White	Negro	White
1 & 2	.011	.031	.04	.18
3	.017	.010	.03	.23
4	.003	.026	.07	.27
5	.019	.158	.21	.22
6	.654	.628	.49	.10
7	.296	.147	.16	.00
Total	1.000	1.000	1.00	1.00

Data in Table 1 indicate that for both the Negro and white sample there was a higher occupational distribution than for the general population. There were more parents in

occupational levels, 3, 4, and 5, and fewer in 6 and 7, than was the case for the general population. This should be considered in any generalizations regarding the sample.

Table 2 illustrates the educational attainment of the parents of the children in the sample, as well as that of the parent population. Here again a discrepancy exists between the sample and the general population as indicated in the 1960 census. Both Negro and white parents have attained a higher level of education than was indicated for the metropolitan area as a whole.

Table 2

Distribution of Educational Attainment of Parent Population and Parents of Children in the Sample

McGuire-White Scale Value	Parent Population Percentage		Sample Population Percentage	
	Negro	White	Negro	White
1 & 2	.029	.067	.06	.10
3	.022	.074	.03	.11
4	.077	.224	.34	.54
5	.148	.224	.27	.17
6	.091	.114	.16	.03
7	.633	.297	.14	.05
Total	1.000	1.000	1.00	1.00

Table 3 contains the means and standard deviations of the sample's chronological age in months by grade and race, data gathered in February and March of the school year. There are approximately the same number of children in each age range in both samples.

The 720 subjects were divided into three reward condition groups of 30 each at each grade level for each race; or, 180 subjects from each grade level.

Table 3

Means and Standard Deviations of Subjects' Chronological Age in Months by Grade and Race

	Age in Months			
	Negro		White	
Grade	Mean	S.D.	Mean	S.D.
2	94.7	6.4	93.8	5.3
4	118.1	7.9	118.5	5.4
7	157.3	8.1	154.9	6.4
10	188.3	7.9	189.3	5.4

Results

The dependent variable for this study, the subjects's reaction time, was converted into means for trials one and two and over-all trial mean and studied through use of a factorial analysis of variance with repeated measures for the variables, reward, trial, grade, intelligence, and sex. Two separate analyses were run because of the number of variables contributing to the between and within subject error estimates, and the limited number of subjects in individual cells: the first analysis of variance took in reward, trial, grade, and intelligence; the second explored reward, trial, grade, and sex. Because of the difficulty in equating IQs for Negro and white subjects, separate analyses were run on the two races and are reported separately.

White Sample

The results of the first analysis of variance run on the white sample data are illustrated in Table 4. Since the major interest of this study was the effect of the verbal incentives, the interaction between praise and blame and the subject variables assumes major importance. However, as illustrated in Table 4, significant relationships among subject variables were noted. The *F*-ratios between grades and between

intelligence levels were significant beyond the .001 level, as was the main effect of the trials.

The interaction between reward condition and trials one and two was also significant beyond the .001 level. Grade level, reward condition, and trials one and two interacted significantly at the .05 level. And the interaction of reward, grade, trial, and intelligence was significant at the .001 level. None of the other variable relationships was significant.

Table 4

White Sample Analysis of Variance Relationships between Reward, Trial, Grade, and Intelligence

Source	df	Sum of Squares	Mean Square	F-ratio
Reward (R)	2	2,201.86	1,100.93	.932
Trial (T)	1	3,622.52	3,622.52	21.812***
Grade (G)	3	104,913.69	34,971.23	29.605***
Intelligence (I)	2	46,860.14	23,430.07	19.835***
R by T	2	12,160.46	6,080.23	36.611***
R by G	6	7,287.24	1,214.54	1.028
R by I	4	7,566.52	1,891.63	1.601
R by T by G	6	2,205.84	367.64	2.214*
R by T by I	4	782.80	195.70	1.178
R by G by I	12	14,548.68	1,212.39	1.026
R by T by G by I	12	4,994.76	416.23	2.506**
T by G	3	470.52	156.84	.944
T by I	2	96.72	48.36	.291
T by G by I	6	1,693.68	282.28	1.700
G by I	6	8,218.26	1,369.71	1.160
Error Between	324	382,721.76	1,181.24	
Error Within	324	53,809.92	166.08	
Total Variation	719	654,155.54		

*P < .05
**P < .01
***P < .001

Illustrated in Table 5 is the relationship between trials, significant at the .001 level. The mean for all subjects on the second trial was .45 seconds faster than the first trial mean. The over-all mean for white subjects was 5.59 seconds.

Table 5
White Sample Mean Reaction Time by Trial

Trial	Mean Reaction Time
One	5.82
Two	5.37
Difference	- .45
Total	5.59

Table 6 illustrates the relationship between grades, significant at the .001 level. Grade seven subjects performed with greatest speed, grade two with the least speed, the difference between these two grades being 2.99 seconds. It would be expected that when performing on the same task, older subjects would finish more quickly. The .15 seconds difference between seventh- and tenth-graders breaks the inverse relationship between reaction time and grade level.

Table 6
White Sample Mean Reaction Time by Grade

Grade	Mean Reaction Time
2	7.64
4	5.29
7	4.65
10	4.80
Total	5.59

Illustrated in Table 7 is the relationship between intelligence levels, significant at the .001 level. As would be expected, subjects of high intelligence completed the task with the greatest speed, the difference between subjects of high and low intelligence being 1.98 seconds. The medium-intelligence subjects performed exactly at the medium reaction time between subjects of high and low intelligence.

Table 7

White Sample Mean Reaction Time by Intelligence

Intelligence[a]	Mean Reaction Time
High	4.61
Medium	5.58
Low	6.59
Total	5.59

[a]High = California Test of Mental Maturity IQ, 110-136; Medium = CTMM IQ, 97-109; Low = CTMM IQ, 60-96.

The relationship between reward condition and subject performance on trials one and two, significant at the .001 level, is illustrated in Table 8. The praise and control groups both decreased their mean reaction time on trial two, the praised group having the greatest decrease. On the other hand, the blamed subjects increased their mean reaction time for trial two.

Table 9 illustrates the relationship between reward condition, grade level, and subject performance on trials one and two, significant at the .05 level. This further breakdown reveals that subjects at all grade levels reacted to praise and no incentive with decreased MRTs for trial two, and to blame with an increased MRT on the second trial, although for second-graders this increase was very slight, .04 seconds. Seventh-graders seemed the most effected by blame.

Table 8

White Sample Mean Reaction Time for Trials One and Two by Reward

Trial	Praise	Blame	Control	Total
One	6.06	5.49	5.91	5.82
Two	4.93	6.19	5.00	5.37
Difference	-1.13	+ .70	- .91	- .45
Total	5.49	5.84	5.45	5.59

At grade levels seven and ten (Table 9), the praised group, although achieving a decreased MRT on trial two, was, even so, behind the no incentive group in magnitude of decrease. For grades two and four praise was more effective than no incentive.

The greatest decrease in mean reaction time occurred for the praised fourth-graders; the greatest increased MRT, for the blamed seventh-graders. Praise was least effective at grade twelve; blame had the least effect at grade two.

Table 10 illustrates the complete breakdown of variable relationship for the first analysis of variance on the data for white subjects. This relationship, significant at the .01 level, includes reward condition, grade level, level of intelligence, and subject performance on trials one and two.

Regardless of intelligence level or grade level, all subjects responded to praise and no incentive with decreased MRTs on the second trial.

The exceptional reaction, evident on this breakdown by intelligence levels, was to the effect of blame. Both low-intelligence second-graders and high-intelligence seventh-graders responded to all incentives with a decreased MRT for trial two. For medium-intelligence fourth-graders blame had no effect, one way or the other. All other subjects, at all levels of intelligence and all grade levels, responded to blame with an increased MRT on the second trial.

Table 9

White Sample Mean Reaction Time for Trials One and Two by Reward and Grade

Grade	Trial	Praise	Blame	Control	Total
2	One	8.51	7.50	7.88	7.96
	Two	7.25	7.54	7.20	7.33
	Difference	-1.26	+ .04	- .68	- .63
	Total	7.88	7.52	7.54	7.64
4	One	6.30	4.95	5.42	5.56
	Two	4.85	5.50	4.76	5.04
	Difference	-1.45	+ .55	- .66	- .47
	Total	5.57	5.22	5.09	5.30
7	One	4.42	4.47	5.35	4.75
	Two	3.45	6.10	4.11	4.55
	Difference	- .97	+1.63	-1.24	- .20
	Total	3.93	5.28	4.73	4.65
10	One	5.03	5.04	5.00	5.01
	Two	4.15	5.65	3.92	4.57
	Difference	- .88	+ .61	-1.08	- .45
	Total	4.59	5.34	4.46	4.79
Total	One	6.06	5.49	5.91	5.82
	Two	4.92	6.20	5.00	5.37
	Difference	-1.13	+ .70	- .91	- .45
	Total	5.49	5.84	5.45	5.59

Praise was most effective for second-graders of medium intelligence (Table 10); least effective for second-graders of high intelligence. Low-intelligence seventh-graders responded

to blame with the greatest increase in mean reaction time.

When looking at levels of intelligence, regardless of grade, both praise and blame had their greatest effect on subjects of low intelligence, their least effect on subjects of high intelligence. All subjects, regardless of intelligence level, responded to praise and no incentive with decreased mean reaction time, and to blame with an increased mean reaction time on the second trial. It is only when grouped according to grade level that the exceptional reaction to blame is evident.

Table 10

White Sample Mean Reaction Time for Trials One and Two by Reward, Grade, and Intelligence

Grade	Intelligence	Trial	Praise	Blame	Control	Total
2	High	One	6.97	7.12	7.43	7.17
		Two	6.48	7.74	6.61	6.94
		Diff.	- .49	+ .62	- .82	- .23
		Total	6.72	7.43	7.02	7.05
	Medium	One	10.10	7.38	7.58	8.35
		Two	7.85	7.73	7.00	7.53
		Diff.	-2.25	+ .35	- .58	- .82
		Total	8.97	7.55	7.29	7.94
	Low	One	8.44	8.01	8.63	8.36
		Two	7.41	7.15	8.00	7.52
		Diff.	-1.03	- .86	- .63	- .84
		Total	7.92	7.58	8.31	7.94
	Total	One	8.51	7.50	7.88	7.96
		Two	7.25	7.54	7.20	7.33
		Diff.	-1.26	+ .04	- .68	- .63

Table 10.--Continued

Grade	Intelligence	Trial	Praise	Blame	Control	Total
2	Total	Total	7.88	7.52	7.54	7.64
4	High	One	4.99	3.87	4.43	4.43
		Two	3.66	5.37	4.00	4.34
		Diff.	-1.33	+1.50	- .43	- .09
		Total	4.32	4.62	4.21	4.38
	Medium	One	6.03	5.23	4.58	5.28
		Two	4.85	5.23	4.02	4.70
		Diff.	-1.18	0.00	- .56	- .58
		Total	5.44	5.23	4.30	4.99
	Low	One	7.87	5.75	7.26	6.96
		Two	6.05	5.91	6.25	6.07
		Diff.	-1.82	+ .16	-1.01	- .89
		Total	6.96	5.83	6.75	6.51
	Total	One	6.30	4.95	5.42	5.56
		Two	4.85	5.50	4.76	5.04
		Diff.	-1.45	+ .55	- .66	- .52
		Total	5.57	5.22	5.09	5.30
7	High	One	4.23	3.03	3.69	3.65
		Two	3.26	2.50	2.83	2.86
		Diff.	- .97	- .53	- .86	- .79
		Total	3.74	2.76	3.26	3.25
	Medium	One	3.93	4.96	4.14	4.34
		Two	3.23	7.12	3.62	4.66

Table 10.--Continued

Grade	Intelligence	Trial	Praise	Blame	Control	Total
7	Medium	Diff.	- .70	+2.16	- .52	+ .32
		Total	3.58	6.04	3.88	4.50
	Low	One	5.09	5.41	8.21	6.24
		Two	3.86	8.66	5.88	6.13
		Diff.	-1.23	+3.25	-2.33	- .10
		Total	4.47	7.04	7.04	6.18
	Total	One	4.42	4.47	5.35	4.75
		Two	3.45	6.09	4.11	4.55
		Diff.	- .97	+1.63	-1.24	- .20
		Total	3.93	5.28	4.73	4.65
10	High	One	4.45	3.51	4.13	4.03
		Two	3.67	3.69	3.27	3.54
		Diff.	- .78	+ .18	- .86	- .49
		Total	4.06	3.60	3.70	3.78
	Medium	One	5.29	5.25	4.77	5.10
		Two	4.66	5.62	3.58	4.62
		Diff.	- .63	+ .37	-1.19	- .48
		Total	4.97	5.43	4.17	4.86
	Low	One	5.34	6.37	6.11	5.94
		Two	4.14	7.65	4.91	5.57
		Diff.	-1.20	+1.28	-1.20	- .37
		Total	4.74	7.01	5.51	5.75

Table 10.--Continued

Grade	Intelligence	Trial	Praise	Blame	Control	Total
10	Total	One	5.03	5.04	5.00	5.02
		Two	4.15	5.65	3.92	4.57
		Diff.	- .88	+ .61	-1.08	- .45
		Total	4.59	5.34	4.46	4.80
Total	High	One	5.16	4.38	4.92	4.82
		Two	4.27	4.82	4.18	4.42
		Diff.	- .89	+ .44	- .74	- .40
		Total	4.71	4.60	4.55	4.61
	Medium	One	6.34	5.70	5.27	5.77
		Two	5.15	6.42	4.55	5.40
		Diff.	-1.19	+ .72	- .72	- .37
		Total	5.74	6.06	4.91	5.58
	Low	One	6.68	6.38	7.55	6.87
		Two	5.36	7.34	6.26	6.32
		Diff.	-1.32	+ .96	-1.29	- .55
		Total	6.02	6.86	6.90	6.59
	Total	One	6.06	5.49	5.91	5.82
		Two	4.92	6.20	5.00	5.37
		Diff.	-1.14	+ .71	- .91	- .45
		Total	5.49	5.84	5.45	5.59

The results of the second analysis of variance run on data for the white sample are illustrated in Table 11. This analysis covers the variables reward, trial, grade, and sex. The

interaction between reward condition, grade level, and sex was significant at the .05 level. Interaction between trials and sex was significant at the .001 level. No other variable relationship interactions were significant.

Table 11

White Sample Analysis of Variance Relationships between Reward, Trial, Grade, and Sex

Source	df	Sum of Squares	Mean Square	F-ratio
Sex (S)	1	.70	.70	.001
R by S	2	3,742.26	1,871.13	1.438
R by T by S	2	317.78	158.89	.929
R by G by S	6	17,289.84	2,881.64	2.215*
R by T by G by S	6	536.58	89.43	.523
T by S	1	2,272.00	2,272.00	13.287***
T by G by S	3	797.34	265.78	1.554
G by S	3	1,530.69	510.23	.392
Error Between	336	437,354.40	1,301.65	
Error Within	336	57,452.64	170.99	
Total Variation	719	654,155.54		

*P < .05
***P < .001

The relationship between reward condition, grade level, and sex, significant at the .05 level, is illustrated in Table 12. The over-all mean reation times show that praised females performed faster than blamed females at every grade level except four, where, for the females, blame was the most effective incentive. For males, however, blame was more effective than praise at two grade levels: two and ten. At the seventh-grade level only, both boys and girls performed more slowly in the blamed group. For other grade levels, the effect of blame varied between the sexes.

Over-all, females performed more quickly in the praised group; males more quickly in the no incentive and blamed groups. It is interesting to note that the over-all means for males and females were only .01 seconds different from each other.

Table 12

White Sample Mean Reaction Time by Reward, Grade, and Sex

Grade	Sex	Praise	Blame	Control	Total
2	Female	6.98	7.88	8.32	7.73
	Male	8.77	7.17	6.77	7.57
	Total	7.87	7.52	7.54	7.65
4	Female	5.78	4.71	4.86	5.12
	Male	5.37	5.74	5.32	5.[illegible]
	Total	5.57	5.22	5.09	5.30
7	Female	4.00	5.18	4.45	4.54
	Male	3.87	5.38	5.00	4.75
	Total	3.93	5.28	4.72	4.64
10	Female	3.97	6.53	4.48	4.99
	Male	5.21	4.17	4.45	4.61
	Total	4.59	5.35	4.46	4.80
Total	Female	5.18	6.07	5.53	5.59
	Male	5.80	5.61	5.38	5.60
	Total	5.49	5.84	5.45	5.59

Table 13 illustrates the performance of white male and female subjects on trials one and two, a relationship significant at the .01 level. Although the females performed faster

than the males on the first trial, and decreased their mean reaction time on trial two, the males still performed .35 seconds faster on trial two. The rate of decrease for mean reaction time between trials one and two was much greater for the male subjects.

Table 13

White Sample Mean Reaction Time for Trials One and Two by Sex

Trial	Female	Male	Total
One	5.64	6.00	5.82
Two	5.55	5.20	5.37
Difference	- .09	- .80	- .45
Total	5.59	5.60	5.59

Summary for White Sample

For white subjects, significant relationships were found:

Between trials.--Over-all mean reaction time from trial one to trial two was reduced.

Between grades.--An inverse relationship was apparent between grade level and over-all mean reaction time, although grade seven performed slightly faster than grade ten.

Between levels of intelligence.--Subjects of high intelligence performed more quickly than subjects of low and medium intelligence.

Between reward condition and performance on trials one and two.--Mean reaction time was decreased from trial one to trial two for subjects in the praise and no incentive groups and increased for subjects in the blame group.

Between reward condition, performance on trials one and two, and grade level.--No difference between grade levels was

apparent in reaction to incentives, although the effect of blame was slight for second-graders and great for seventh-graders.

Between reward condition, performance on trials one and two, grade level, and level of intelligence.--All subjects, regardless of grade level or level of intelligence, reacted to praise and no incentive with decreased mean reaction time on trial two. Most subjects, regardless of grade level or level of intelligence, reacted to blame with increased mean reaction time on trial two: the exceptions were low-intelligence second-graders and high-intelligence seventh-graders, who decreased their mean reaction time on trial two regardless of incentive; and medium-intelligence fourth-graders, for whom there was no change in mean reaction time from trial one to trial two under the blame incentive.

Between reward condition, grade level, and sex.--Only at the seventh-grade level did both sexes respond to praise and blame in the same manner: for both males and females praise was the more effective incentive at the seventh-grade level. At all other grade levels the effect of the incentives varied between the sexes: in general, females responded more favorably to praise, males to blame.

Between performance on trials one and two, grade level, and sex.--White male subjects reduced their mean reaction time on trial two to a greater extent than did female subjects.

Negro Sample

Data for the Negro subjects were analyzed exactly as for white subjects. The first analysis of variance, including the variables reward, trial, grade, and intelligence, is illustrated in Table 14. The relationships between trials, between levels of intelligence, between reward conditions and trials, and between reward conditions, trials, and grade level all were significant for the Negro sample.

Table 14

Negro Sample Analysis of Variance Relationships between Reward, Trial, Grade, and Intelligence

Source		df	Sum of Squares	Mean Square	F-ratio
Reward	(R)	2	5,284.68	2,642.34	1.012
Trial	(T)	1	5,951.84	5,951.84	12.242***
Grade	(G)	3	8,434.41	2,811.47	1.077
Intelligence	(I)	2	24,825.10	12,412.50	4.756**
R by T		2	16,712.52	8,356.26	17.188***
R by G		6	20,384.04	3,397.34	1.302
R by I		4	4,101.16	1,025.29	.393
R by T by G		6	8,131.44	1,355.24	2.788*
R by T by I		4	1,848.72	462.18	.951
R by G by I		12	29,550.60	2,462.55	.944
R by T by G by I		12	1,938.00	161.50	.332
T by G		3	1,285.11	428.37	.881
T by I		2	450.36	225.18	.463
T by G by I		6	2,904.12	484.02	.996
G by I		6	29,357.04	4,892.84	1.875
Error Between		324	845,581.68	2,609.82	
Error Within		324	157,519.08	486.17	
Total Variation		719	164,259.96		

*P < .05
**P < .01
***P < .001

The relationship between trials, significant at the .001 level, is illustrated in Table 15. The mean reaction time was reduced by .58 seconds from trial one to trial two. The overall mean reaction time for Negro subjects was 7.50 seconds.

Table 15

Negro Sample Mean Reaction Time by Trial

Trial	Mean Reaction Time
One	7.79
Two	7.21
Difference	- .58
Total	7.50

Table 16 illustrates the relationship for the Negro sample between intelligence levels, significant at the .01 level. The over-all mean for the medium-intelligence group was faster than that of even the high-intelligence group by .29 seconds, an unusual finding.

Table 16

Negro Sample Mean Reaction Time by Intelligence

Intelligence[a]	Mean Reaction Time
High	7.24
Medium	6.95
Low	8.31
Total	7.50

[a]High = California Test Mental Maturity IQ, 89-132; Medium = CTMM IQ, 78-88; Low = CTMM IQ, 60-77.

The relationship between reward condition and subject performance on trials one and two, significant at the .001 level, is illustrated in Table 17. The praise and control groups both decreased their mean reaction time on trial two, the praised group having the greatest decrease. On the other hand,

the blamed subjects increased their mean reaction time for trial two.

Table 17

Negro Scample Mean Reaction Time for Trials One and Two by Reward

Trial	Praise	Blame	Control	Total
One	8.03	7.50	7.83	7.79
Two	6.60	8.27	6.77	7.21
Difference	-1.43	+ .77	-1.06	- .58
Total	7.31	7.88	7.30	7.50

Table 18 illustrates the relationship between reward condition, grade level, and performance on trials one and two, significant at the .05 level. Subjects in the praised group, at all grade levels, responded with a decreased MRT for trial two.

Second-graders reacted exceptionally to blame: they achieved a greater decrease in mean reaction time for trial two with blame than with praise. In fact, for Negro second-graders, no incentive was more effective than either praise or blame. Subjects at all other grade levels reacted to blame with an increased mean reaction time on the second trial.

For the no incentive group, it was the fourth-graders who were the exception: they reacted to no incentive with an increased mean reaction time, the only subject grouping to do so.

Praise was most effect at the tenth-grade level, least effective for second-graders. Blame had its greatest effect at the seventh-grade level

In Table 18 it can be seen that the inverse relationship between grade level and mean reaction time, apparent for the white sample, was not evident for the Negro sample. The second- and seventh-graders performed almost equally rapidly, with fourth-graders performing more slowly than any other grade level.

Table 18

Negro Sample Mean Reaction Time for Trials One and Two by Reward and Grade

Grade	Trial	Praise	Blame	Control	Total
2	One	7.79	7.42	7.38	7.53
	Two	7.26	6.76	6.25	6.76
	Difference	- .53	- .66	-1.13	- .77
	Total	7.52	7.09	6.81	7.14
4	One	8.93	7.02	8.01	7.98
	Two	7.46	7.95	8.14	7.85
	Difference	-1.47	+ .93	+ .13	- .13
	Total	8.19	7.48	8.07	7.91
7	One	6.89	7.50	8.03	7.47
	Two	5.31	8.96	6.40	6.89
	Difference	-1.58	+1.46	-1.63	- .25
	Total	6.10	8.23	7.21	7.18
10	One	8.50	8.05	7.91	8.15
	Two	6.37	9.40	6.28	7.35
	Difference	-2.13	+1.35	-1.63	- .80
	Total	7.43	8.72	7.09	7.75
Total	One	8.03	7.50	7.83	7.79
	Two	6.60	8.27	6.77	7.21
	Difference	-1.43	+ .77	-1.06	- .58
	Total	7.31	7.88	7.30	7.50

The results of the second analysis of variance run on data for Negro subjects, using the variables reward, grade, trial,

and sex, are illustrated in Table 19. Sex does not significantly react with any variable for Negro subjects.

Table 19

Negro Sample Analysis of Variance Relationships between Reward, Trial, Grade, and Sex

Source	df	Sum of Squares	Mean Square	F-ratio
Sex (S)	1	203.40	203.40	.077
R by S	2	2,143.32	1,071.66	.407
R by T by S	2	282.12	141.06	.295
R by G by S	6	1,447.56	241.26	.092
R by T by G by S	6	2,418.84	403.14	.842
T by S	1	562.36	562.36	1.174
T by G by S	3	385.29	128.43	.268
G by S	3	20,263.47	6,754.49	2.568
Error Between	336	883,817.76	2,630.41	
Error Within	336	160,893.60	478.85	
Total Variation	719	142,386.60		

Summary for Negro Sample

For Negro subjects, significant relationships were found:

Between trials.--Over-all mean reaction time from trial one to trial two was reduced.

Between levels of intelligence.--Subjects of medium intelligence performed more quickly than subjects of high and low intelligence.

Between reward condition and performance on trials one and two.--Mean reaction time was decreased from trial one to trial two for subjects in the praise and no incentive groups and increased for subjects in the blame group.

Between reward condition, performance on trials one and two, and grade level.--Negro second-graders decreased mean

reaction time on trial two regardless of incentive, and fourth-graders decreased mean reaction time on trial two only for praise. At the seventh- and tenth-grade levels, subject response to incentives was the same as for the population as a whole. All Negro subjects, at all grade levels, responded to praise with decreased mean reaction time.

Over-All Summary

Significant relationships found in this study were:

Between trials.--For both Negro and white subjects the over-all mean reaction time from trial one to trial two was reduced.

Between grades.--This relationship was significant only for the white subjects, where an inverse relationship was noticed between grade level and over-all mean reaction time, although grade seven performed slightly faster than grade ten.

Between levels of intelligence.--This relationship was significant for both white and Negro subjects, but it was different for each race: for white subjects the high-intelligence group performed more quickly; for Negro subjects the medium-intelligence group performed more quickly.

Between reward condition and performance on trials one and two.--For both Negro and white subjects, the mean reaction time was decreased from trial one to trial two for subjects in the praise and no incentive groups and increased for subjects in the blame group.

Between reward condition, performance on trials one and two, and grade level.--For white subjects no difference between grade levels was apparent in reaction to incentives, although the effect of blame was slight for second-graders and great for seventh-graders. For Negro subjects, both second- and fourth-graders reacted exceptionally to the incentives: second-graders decreased mean reaction time for trial two regardless of incantive; fourth-graders decreased mean reaction time only for praise. All subjects, regardless of race or grade level,

responded to praise with decreased mean reaction time on trial two.

Between reward condition, performance on trials one and two, grade level, and level of intelligence.--This relationship was significant only for the white subjects, where all subjects, regardless of grade level or intelligence level, reacted to praise with decreased mean reaction time on trial two. Low-intelligence second-graders, high-intelligence seventh-graders, and medium-intelligence fourth-graders reacted to blame exceptionally: the second- and seventh-grade groups decreased mean reaction time on trial two regardless of incentive; and the fourth-grade group did not react to blame one way or the other; that is, the mean reaction time for both trials one and two was the same.

Between reward condition, grade level, and sex.--This relationship was significant for only the white sample, where at only the seventh-grade level did both sexes react to the incentives as did the study subjects as a whole. At the other grade levels the reaction of the sexes was varied: in general, females responded more favorably to praise, males to blame.

Between performance on trials one and two, grade level, and sex.--This relationship was also only significant for the white sample, where the males reduced their mean reaction time on trial two to a greater degree than did female subjects.

Analyses of variance between schools, between examiners, and between socio-economic levels were non-significant for both white and Negro subjects. None of these variables had any significant effect on the reaction of subjects, regardless of race, grade level, level of intelligence, or sex, to the verbal incentives under study.

Conclusions

In terms of the praise and blame condition established for this study, that of immediate and private feedback, after a trial of 32 perceptual speed responses, as to the adequacy of

the subject's performance in relation to the expectancy of the examiner, several significant relationships were observed.

All subjects, regardless of sex, race, grade level, or level of intelligence, reacted to praise with decreased mean reaction time from trial one to trial two.

All subjects, regardless of sex, race, grade level, or level of intelligence, reacted to no incentive with decreased mean reaction time from trial one to trial two, with the one exception of Negro fourth-graders, who slightly increased their mean reaction time on trial two under no incentive.

All subjects, regardless of sex, race, grade level, or level of intelligence, responded to blame with increased mean reaction time on the second trial, with the exception of: low-intelligence white second-graders, high-intelligence white seventh-graders, and second-grade Negro subjects, who decreased mean reaction time on trial two regardless of incentive; and medium-intelligence white fourth-graders, who had no change in performance from trial one to trial two under the condition of blame.

These exceptions are among the interesting findings of this study and an adequate interpretation is difficult. A likely explanation for the exceptional reaction to blame of the second-graders is their lack of understanding of the implications of reproof. For the high-intelligence seventh-graders, this explanation is not so likely. Reproof had such devastating effect on medium- and low-intelligence seventh-graders that one is inclined to feel that the "second-chance" offer was understood only by those of high-intelligence. Yet this is not a satisfactory answer either, since it doesn't hold true for other grade levels.

It is interesting to note that, due to the method of arriving at the levels of intelligence in this study, the low-intelligence range for white subjects encompasses the low- and medium-intelligence ranges for Negro subjects, and the lower IQs in the Negro high-intelligence range. This point has interesting implications for further research.

For the white sample, an inverse relationship was found between grade level and mean reaction time, although the seventh-graders did perform slightly faster than tenth-graders. For Negro subjects, tenth-graders worked at the same rate as fourth-graders, seventh-graders at the same rate as second-graders, approximately. At first glance, it might be felt that this is just cause to doubt the validity of the Negro sample. However, since almost exactly the same reactions to incentives were found in both races, it appears that the Negro subjects were responding differentially to the incentives, which was the major interest of the study.

When using a criterion measure such as the one in this study, the possibility of a physiological limit effecting the amount of change must be considered. For white subjects, it was noticed that males had a greater decreased mean reaction time from trial one to trial two than did females, although the females performed more quickly on trial one than did the males. It might be argued that the female subjects had less "room" for change than did the male subjects, and thus the magnitude of change is not a function of performance ability, but of physiological limit. However, the mean reaction times found in this data analysis in no way approach the lower limits of observed reaction time on the task. Reactions to individual stimulus cards as low as one or two seconds are not at all uncommon. The writer, therefore, feels that there is sufficient evidence to discount the probability of physiological limits having significantly effected the obtained data.

Probably the most striking finding of this study was the consistency of the effect of praise and blame. The study grew out of the observation from the literature that the effects of praise and blame were extremely variable and highly reliant on subject and situational modifications. The results of this study indicate that the effects of praise and blame are quite obvious and quite consistent, irrespective of some rather basic differences in subject variables.

Perhaps this consistency of effects resulted from the task used being relatively uneffected by variables other than verbal incentive, and thus, unlike other tasks in other studies, the results reflected only changes resulting from the incentive.

These results by no means answer any questions with regard to situational variables, but perhaps they offer a point of reference, or a baseline, from which to work in approaching an answer to the question of the effectiveness of praise and reproof in variable conditions.

Due to basic differences in dependent measures, the results of this study cannot be compared adequately to others in the literature, and thus no statements can be made as to their agreement or disagreement. They may, however, afford a framework within which to compare more meaningfully the results of many studies dealing with praise and reproof.

In any event, in situations involving perceptual motor skills, with speed as the criterion measure, these results should hold. That is, reproof should have a debilitating effect upon performance, while praise results in a somewhat higher increase in performance than that coming from practice alone.

Three points for further research would seem most relevant now. First, what would be the result of using Negro examiners for the Negro subjects, instead of white examiners? Kennedy and Vega are currently following this lead. The second, and much more difficult in design, is the study of the satiation effect of these reward conditions over time. Third, and somewhat more obvious follow-up, would be a replication of this study using a task more similar to school activities. The review of the literature has pointed up the hazards of this procedure, mainly in the area of equating experience.

Summary

Experimental studies using various rewards and punishments for motivating discrimination, learning behavior, and performance in human and animal subjects compose the greater part

of the literature of psychology. The history of child psychology is spanned by studies seeking to evaluate the effectiveness of praise and blame as incentives for school children, a literature sufficiently rich, broad, and confusing in and of itself to warrant special attention.

In looking over the studies prior to 1930, confusion of results is already apparent. In the two earlier studies, verbal encouragement was found to improve the scores of all participating children on both the motor task of Binet & Vaschide (1897) and the regular classroom task of Kirby (1913). Neither study employed a control group or attempted to compare effects of praise versus reproof.

A very careful researcher, Hurlock found, in two studies (1924, 1925a), one more or less a duplication of the other, neither praise or reproof to be superior, although both were more effective than practice. In a third study (1925b), she found praise to be decidedly more effective than reproof, although both were still superior to practice. This latter study was substantiated by Cohen (1927) in a duplicate study resulting in the same conclusions. Hurlock's first two studies used an intelligence test as the task, while the latter study made use of a regular classroom task.

Bird (1927), on a reading readiness test administered to four- and six-year-olds, found that for some children withheld praise caused delayed learning and that for some children reproof was needed to encourage their best efforts.

None of these studies with differing results used the same age levels: Kirby used third- and fourth-grades; Hurlock's two earlier studies used grades three, five, and eight, while her later study, and that of Cohen, used grades four and six; Bird used nursery and kindergarten children. Yet in the studies using multiple grades, no age differences were found. Could differences in methodology be responsible for such clashing differences in results?

On one point all the studies in the 1930-1940 decade agree: the experimental incentives were administered regardless

of the actual performance of the children. This in itself is a somewhat artificial approach, since classroom teachers would supposedly be using the incentives to fit the occasion. But whether or not the answers were correct or incorrect, or the performance satisfactory or unsatisfactory, during this decade, as for the previous pertinent research, subjects were administered the incentives at random.

Warden & Cohen (1931) not only found praise and reproof to be not significantly different from each other or the control group, but also to be inferior to promise of a game, in motivating fourth-graders in a regular classroom task. Brenner (1934) also found no differences between praise, blame, and control groups, even when divided into delayed and immediate praise and blame. Nor did Benton (1936) find any differences between praise and control: he did not make use of a reproved group.

Yet Chase (1931), using a motor task, found both praise and reproof to be better than the control, and failure to be better even than praise. However, this latter finding was contaminated by previous incentive condition and method of administering the task. Anderson & Smith (1933), in a follow-up study, agreed with Chase's finding and definitely concluded blame to be more effective than praise. And Forlano & Axelrod (1937), endeavoring to take into consideration the personality of the subjects, found a significant difference only in the introverted subjects, where blame was more effective than praise or control. There were no differences in effectiveness of incentives for the extroverts.

None of these later studies agreed with Hurlock (1925b), that praise is more effective than reproof. The tendency for this period of research leans either to no differences between praise and reproof, or reproof to be more effective. There is still no generally accepted conclusion on the effectiveness of verbal incentives for school children. It is no wonder that researchers qualified their reports with statements to the effect

that praise and blame must be used wisely and with understanding for the specific pupil at the specific time.

Whether it was the effect of World War II, a shift in interest by investigators in the field, or high standards for the next step set by Schmidt (1941), the decade 1940-1950 added only three studies to the investigation of the effects of praise and blame on the performance of school children. Only one of these, Thompson & Hunnicutt (1944), endeavored to compare the effects of both praise and blame on student subjects.

Potter (1943) found that reproof decreased performance of third-graders, increased that of sixth-graders, and had little effect on ninth- and twelfth-graders.

In an effort to establish a relationship between praise versus a money reward as incentives, Klugman (1944) found no differences in performance on the Stanford-Binet. His is the first study since that of Bird (1927) not to employ a control group.

In their replication of the Forlano & Axelrod (1937) study on the effects of praise and blame on introverts and extroverts, Thompson & Hunnicutt (1944) discovered that adding three more trials to the original three trials changed the significant results of the study. Forlano & Axelrod could only conclude that blame was better for introverts. On the basis of repeated praise and blame, Thompson & Hunnicutt concluded praise to be better for introverts and blame to be better for extroverts.

The decade 1950-1960 doubled the number of studies of the previous decade, and introduced interesting variations to the research theme. Blind children and mental defectives were studied (Kent, 1956; Ellis & Distephano, 1958). The experimental incentives were administered according to the performance of the subjects (Terrell & Kennedy, 1958). A localization task was used (Sandstrom & Weinz, 1958), and massed versus spaced praise was studied (Mech, Kapos, Hurst, & Auble, 1954). Teacher use of praise or reproof in the classroom was recorded in an effort to evaluate its effect on reading improvement (Silverman, 1957).

Results were just as varied, but almost predictable: massed praise is more effective than spaced praise; even for blind children praise and blame are more effective than control, except for public-school blind where there is no difference between groups and where performance itself is significantly inferior to that of residential-school blind and public-school sighted children; teacher use of praise and reproof in a classroom cannot predict reading growth in pupils; candy is a significantly more effective reward than praise or reproof, neither of which is better than a knowledge-of-results control group; praise is significantly more effective for mental defectives than no incentive; reproof is better than praise if the reward offered is unattainable, although both are better than no information on results of performance.

Although one study found reproof to be more effective, the conclusion was qualified by experimental conditions in the praised group. No really radical findings are noted for this decade. Possibly its importance lies in the varied forms that the research was taking, in an ever-growing effort to try to reach some valid decision as to the effects of praise and reproof on the performance of school children.

In the early 1960s five of the seven studies pertain directly or indirectly to the problem of establishing a relationship between verbal incentives and intelligence: Stevenson & Snyder, 1960; Kennedy, Turner, & Lindner, 1962; Metz, 1962; Tiber & Kennedy, 1963; Willcutt & Kennedy, 1963.

Stevenson & Snyder, with a more complicated research design than that of Ellis & Distephano (1958), found that for mentally-retarded children, changing the incentives changed their effect on the subjects: praise as the original incentive maintained higher performance even when followed by blame or no incentive, while with more praise performance was increased; blame followed by praise led to increased performance, but a decrement in performance was apparent for blame followed by no incentive.

Metz, incidentally studying effects of incentives on intelligence, found both stress (subjects told they had failed) and praise to facilitate the creative performance of inferior students. For superior students creative performance was inhibited by stress, while either praise or neutral instructions led them to superior performance.

Two studies in this decade illustrated a difference in the interaction between verbal incentive and intelligence by using the same discrimination task and procedure at different grade levels. For eleventh- and twelfth-graders Kennedy, Turner, & Lindner found no difference between the effects of praise, blame, or no incentive on the superior students, yet for students of average ability in the reproved group, a significant decrement in performance was obvious. Willcutt & Kennedy found that for all fourth-graders, regardless of intelligence level, praise was the most effective incentive.

Changing the task from a discrimination problem to performance on an intelligence test, the grade level to second and third, and controlling for class and caste social groups rather than level of intelligence, Tiber & Kennedy, contrary to expectations, found no difference in effectiveness of incentives for any of the social groups, yet found a significant difference in IQ scores between these social groups. The question arises: was it the difference in task, the difference in grade level, or the confounding of intelligence with social groupings that was responsible for these results?

The study of Dollins, Angelino, & Mech (1960), which concluded that teacher use of praise in the elementary classroom did not effect student personality, or performance on an arithmetic task, also raises an interesting question. Could the finding of Mech, Kapos, Hurst, & Auble (1954), that massed praise is more effective than spaced praise, have a bearing on the discrepancy between the Dollins-Angelino-Mech study and the Willcutt-Kennedy study? The first study was spaced over 30 days, with praise at its maximum given on only 20 of those days,

while the Willcutt-Kennedy study, for the individual child, was completed in only a short time on one day, with the praise being given immediately for a specific task at hand.

The results of the seventh study in the early 1960s are in agreement with earlier research: Fischer (1962) found, just as did Terrell & Kennedy (1957), that an immediate food reward is the most effective incentive for youngsters.

These studies seem to suggest that in some way there is an interaction between intelligence, age, even social group, and the effects of praise and blame on school children. But in looking back over these studies, the most apparent problem is that although results and conclusions are different and diversified, so are problems, and methods, and subjects, and tasks, and method of statistical analysis, and factors controlled for and factors not considered, and so on almost ad infinitum. Even after 65 years and 29 studies endeavoring to shed light on the relative effectiveness of incentives for school children, it can not be said with authority: "This is so, for this particular child in this particular grade at this particular level of intelligence in this particular social group." The definitive study has not yet been done.

Because of the very nature of the classroom setting there is little question that verbal praise and blame are the most often used incentives in the armamentaria of the teacher. These incentives have a history of long use, as well as a large literature on both a philosophical and empirical basis. Often, however, the incentives have been considered one at a time, and even in studies where both were considered, the comparability of findings is open to question due to the lack of uniformity in methodology and conditions. What seems to be needed now is a study which attempts to take into account those variables which are known, or have been suggested, to have a bearing on the effectivenss of verbal incentives with school children.

The present study was an investigation of the effects of praise and blame on a discrimination task under the variables of

grade, intelligence, sex, race, social class, school, and examiner. Discrimination and learning were selected as the dependent variable because of the relevance to a classroom setting. Although all of the subjects could solve the task used in this study, the speed with which it was solved seemed highly variable. The analysis of this variance was the major statistic of the study.

The 720 subjects were drawn from a large southern metropolitan area and divided into three reward conditions: praise, blame, no incentive--240 subjects each; four grade levels: two, four, seven, ten--180 subjects each; three levels of intelligence: high, medium, low--240 subjects each; and two races: Negro, white--360 subjects each. There were 5 subjects in each of 144 cells.

Four instruments were used in the investigation: the California Test of Mental Maturity to establish intelligence level, a Wollensak tape recorder to monitor all tests, an Esterline Angus 20-pen event recorder to record all subject and examiner responses, and a four-key discrimination box designed for the study.

The 32 oddity-problem stimulus cards, identical to those used by Kennedy, Turner, & Lindner (1962) and Willcutt & Kennedy (1963), presented four patterns, one of which was different from the other three. The problem was to identify the odd pattern as quickly as possible by depressing the correct key on the discrimination box. The examiner was as active as necessary to assure that the subjects fully understood what was required of them.

Data was simultaneously gathered for grades two, four, seven, and ten, all tests being administered in private in a quiet room at the school of each subject. Subjects were administered the 32 stimulus cards, followed by the experimental reward condition, and a second trial on the same stimulus cards.

The reaction time between the stimulus cards appearing on the viewing screen and the depression of the subject key was

the criterion measure. The major variance was the mean reaction time of trials one and two, the over-all mean reaction time for both trials, and the mean reaction time difference between trials one and two.

Lindquist repeated measures analysis of variance was used to analyze data in terms of reward condition, grade level, level of intelligence, and sex. In addition to these variables, the data were studied in relation to soci-economic level, examiner variability, and variability between schools used in the study.

Several significant relationships were found. For both Negro and white subjects the over-all mean reaction time from trial one to trial two was reduced. For white subjects there was an inverse relationship apparent between grade level and mean reaction time, a relationship not apparent in the Negro sample. White subjects of high intelligence performed the most quickly; while for the Negro subjects, the medium-intelligence group performed most quickly.

For both Negro and white subjects the mean reaction time from trial one to trial two was decreased under the conditions of praise and no incentive and was increased under the condition of blame.

White subjects at all grade levels reacted to the reward conditions in the same manner, although the effect of blame was slight for second-graders. For the Negro sample, second-graders decreased their mean reaction time on trial two regardless of incentive, fourth-graders decreased it only under the condition of praise. All subjects, regardless of race or grade level, responded to praise with decreased mean reaction time on trial two.

For white subjects there was a significant relationship between grade level reaction and level of intelligence. Both low-intelligence second-graders and high-intelligence seventh-graders reacted to all incentives with decreased mean reaction time on trial two.

For white subjects there were significant differences between the sexes. Girls responded more favorably to praise, boys to blame. And the males reduced their mean reaction time on trial two to a greater degree than did the females. It is felt that due to the mean reaction times used in this study being considerably higher than the lower limits of observed reaction time, that the probability of physiological limits having significantly effected the obtained data can be discounted.

The effects of examiner variability, socio-economic level, or variability among schools were found to be non-significant to the results of this study.

Probably the most striking finding was the consistency of the effects of praise and blame. The study grew out of the observation from the literature that the effects of praise and blame were extremely variable and highly reliant on subject and situational modifications. The results of this study indicate that the effects of praise and blame are quite obvious and quite consistent, irrespective of some rather basic differences in subject variables.

Perhaps this consistency of effect resulted from the task used being relatively uneffected by variables other than verbal incentive, and thus, unlike other tasks in other studies, the results reflected only changes resulting from the incentive.

These results, however, by no means answer any questions with regard to the situational variable. But perhaps they do offer a point of reference from which to work in approaching an answer to the question of the effectiveness of praise and reproof in variable conditions.

In any event, in situations involving perceptual motor skills, with speed as the criterion measure, these results should hold. That is, reproof should have a debilitating effect upon performance, while praise results in a somewhat higher increase in performance than that coming from practice alone.

References

Anderson, H. H., & Smith, Ruth S. Motivation of young children: the constancy of certain behavior patterns. J. exp. Educ., 1933, 2, 138-160.

Benton, A. L. Influence of incentives upon intelligence test scores of school children. J. genet. Psychol., 1936, 49, 494-497.

Binet, A., & Vaschide, E. Experiences de forces musculaire et de fond chez les jeunes garcons. Annee Psychol., 1897, 4, 15-63.

Bird, Grace E. Personality factors in learning. Personnel J., 1927, 6, 56-59.

Bluhm, P. M. Discrimination reaction time as a function of reinforcement-related DRQ anxiety and task difficulty. Unpublished Ph.D. dissertation, Florida State University, 1963.

Brenner, B. Effect of immediate and delayed praise and blame upon learning and recall. Teach. Coll. Contrib. Educ., 1934, No. 620.

Briggs, T. H. Sarcasm. Sch. Rev., 1928, 36, 685-695.

Chase, Lucille E. Motivation of young children: an experimental study of the influence of certain types of external incentives upon the performance of a task. Univ. Iowa Stud., Child Welf., 1932, 5, No. 3.

Cohen, A. A study of incentives under schoolroom conditions. Unpublished M.A. thesis, Columbia University, 1927.

Davis, R. A., & Ballard, C. R. Effectiveness of various types of classroom incentives. Educ. Meth., 1932, 12, 134-145.

Dollins, J. G., Angelino, H., & Mech, E. V. With words of praise. Elem. Sch. J., 1960, 446-450.

Ellis, N. R., & Distephano, M. K. Effects of verbal urging and praise upon rotary pursuit performance in mental defectives. Amer. J. ment. Defic., 1959, 64, 486-490.

Fischer, W. F. Sharing in preschool children as a function of amount and type of reinforcement. Unpublished Ph.D. dissertation, The University of Connecticut, 1961.

Forlano, H., & Axelrod, C. The effect of repeated praise or blame on the performance of introverts and extroverts. J. educ. Psychol., 1937, 28, 92-100.

Gates, Georgina S., and Rissland, Louise Q. The effect of encouragement and of discouragement upon performance. J. educ. Psychol., 1923, 14, 21-26.

Gilchrist, E. P. The extent to which praise and reproof affect a pupil's work. Sch. and Soc., 1916, 4, 872-874.

Hurlock, Elizabeth B. The value of praise and reproof as incentives for children. Arch. Psychol., 1924, No. 71.

Hurlock, Elizabeth B. The effects of incentives upon the constancy of the I.Q. J. genet. Psychol., 1925, 32, 422-434. (a)

Hurlock, Elizabeth B. An evaluation of certain incentives used in school work. J. educ., Psychol., 1925, 16, 145-159. (b)

Hurlock, Elizabeth B. The pscyhology of incentives. J. soc. Psychol., 1931, 2, 261-290.

Kennedy, W. A., Turner, A. J., & Lindner, R. Effectiveness of praise and blame as a function of intelligence. Percept. mot. Skills, 1962, 15, 143-149.

Kent, H. R. The effect of repeated praise and blame on the work achievement of blind children. Unpublished Ph.D. dissertation, Columbia University, 1956.

Kirby, T. J. Practice in the case of school children. Teach. Coll. Contrib. Educ., 1913, No. 58.

Klugman, S. F. The effect of money incentive versus praise upon the reliability and obtained scores of the revised Stanford-Binet Test. J. general Psychol., 1944, 30, 255-269.

Mech, E., Kapos, E., Hurst, F., & Auble, D. A mathematical description of the relationship between frequency of responding and intertrial interval in a motivational situation. J. Psychol., 1954, 37, 251-256.

Metz, W. A. The relative effects of stress and praise on creativity. Unpublished Ph.D. dissertation, Florida State University, 1961.

Murphy, G., Murphy, Lois B., & Newcomb, T. M. Experimental social psychology. New York: Harper & Bros., 1937. Chapt. VII. Characteristic social behavior of children in our culture: aggression and competition. Pp. 376-508.

Potter, E. H. The effects of reproof in relation to age in school children. J. genet. Psychol., 1943, 63, 247-258.

Sandstrom, C. I., & Weinz, E. Effects of praise and reproof in a localization experiment. Acta. Psychol., 1958, 14, 137-143.

Schmidt, H. O. The effects of praise and blame as incentives to learning. Psychol. Monog., 1941, 53, No. 3.

Silverman, H. F. Effects of praise and reproof on reading growth in a non-laboratory classroom setting. J. educ. Psychol., 1957, 48, 199-206.

Stevenson, H. W., & Snyder, Leila C. Performance as a function of the interaction of incentive conditions. J. Pers., 1960, 28, 1-11.

Terrell, G., Jr., & Kennedy, W. A. Discrimination learning and transposition in children as a function of the nature of the reward. J. exper. Psychol., 1957, 53, 257-260.

Thompson, G. G., & Hunnicutt, C. W. Effects of repeated praise or blame on the work achievements of introverts and extroverts. J. educ. Psychol., 1944, 35, 257-266.

Tiber, N., & Kennedy, W. A. The effects of incentives on the intelligence test performance of different social groups. J. consult. Psychol., in press.

Turner, A. J. Discrimination reaction time as a function of anxiety and task difficulty. Unpublished Ph.D. dissertation, Florida State University, 1962.

Warden, C. J., & Cohen, A. A study of certain incentives applied under schoolroom conditions. J. genet. Psychol., 1931, 39, 320-327.

Willcutt, H. C., & Kennedy, W. A. Relation of intelligence to effectiveness of praise and reproof as reinforcers for fourth-graders. Percept. mot. Skills, in press.

Fig. 4.--Stimulus Cards at Difficulty Level 1

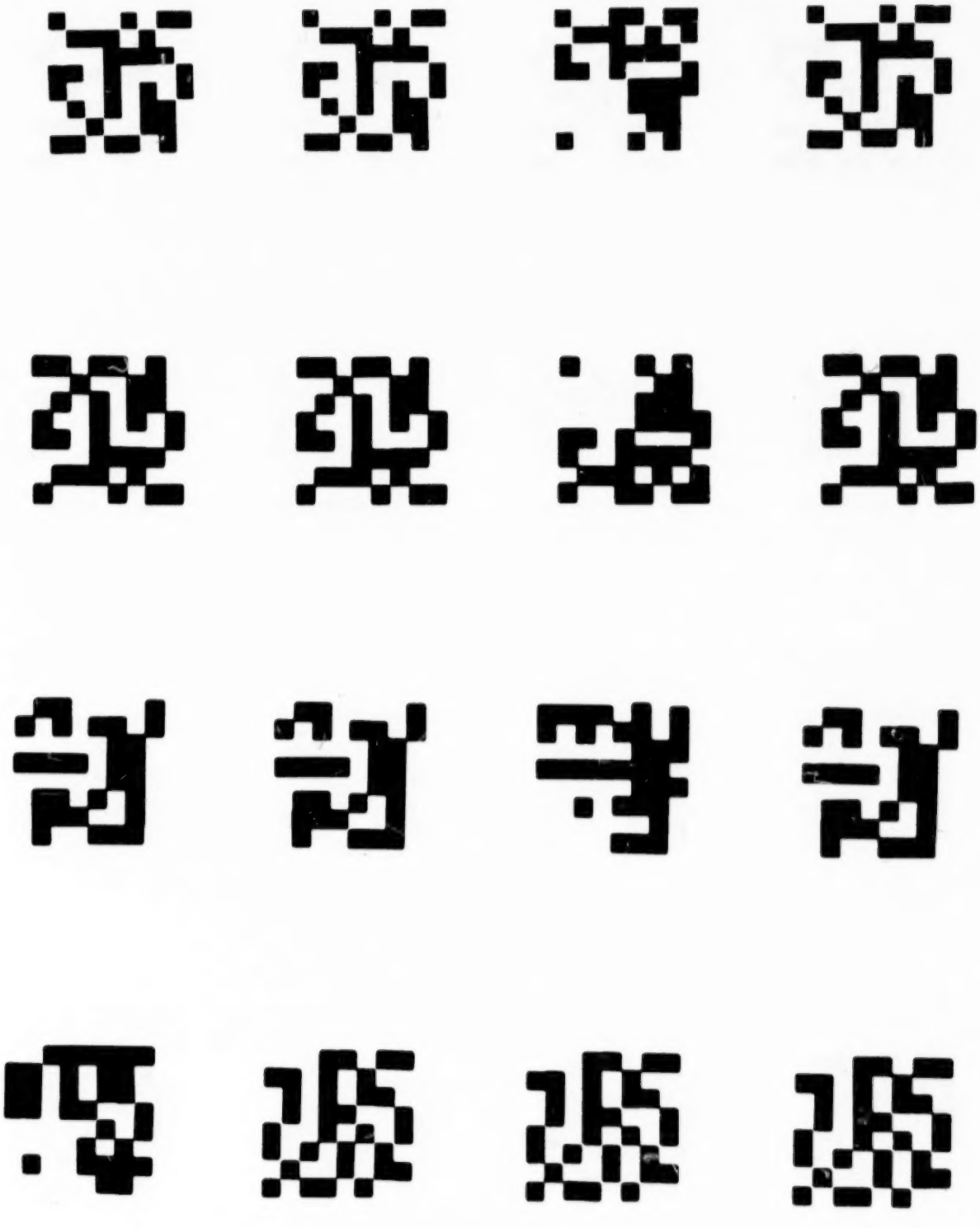

Fig. 5.--Stimulus Cards at Difficulty Level 2

Fig. 6.--Stimulus Cards at Difficulty Level 3

Fig. 7.--Stimulus Cards at Difficulty Level 4

Fig. 8.--Stimulus Cards at Difficulty Level 5

Fig. 9.--Stimulus Cards at Difficulty Level 6

Fig. 10.--Stimulus Cards at Difficulty Level 7

Fig. 11.--Stimulus Cards at Difficulty Level 8